Copyright Page

Edited by Douglas W. Portmore

The Oxford Handbook of Consequentialism

Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy Online Publication Date: Oct 2020

(p. iv) Copyright Page

Oxford University Press is a department of the University of Oxford. It furthersthe University's objective of excellence in research, scholarship, and educationby publishing worldwide. Oxford is a registered trade mark of Oxford UniversityPress in the UK and certain other countries.

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America.

© Oxford University Press 2020

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of Oxford University Press, or as expressly permitted by law, by license, or under terms agreed with the appropriate reproduction rights organization. Inquiries concerning reproduction outside the scope of the above should be sent to the Rights Department, Oxford University Press, at the address above.

You must not circulate this work in any other form and you must impose this same condition on any acquirer.

Library of Congress Cataloging-in-Publication DataNames: Portmore, Douglas W., editor. Title: The Oxford handbook of consequentialism / Douglas W. Portmore. Description: New York : Oxford University Press, 2020. | Summary: "This handbook contains thirty-two previously unpublished contributions to consequentialist ethics by leading scholars, covering what's happening in the field today as well as pointing to new directions for future research. Consequentialism is a rival to such moral theories as deontology, contractualism, and virtue ethics. But it's more than just one rival among many, for every plausible moral theory must concede that

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

the goodness of an act’s consequences is something that matters even if it’s not the only thing that matters. Thus, all plausible moral theories will accept both that the fact that an act would produce good consequences constitutes a moral reason to perform it and that the better that act’s consequences the moral reason there is to perform it. Now, if this is correct, then much of the research concerning consequentialist ethics is important for ethics in general. For instance, one thing that consequentialist researchers have investigated is what sorts of consequences matter: the consequences that some act would have or the consequences that it could have-if, say, the agent were to follow up by performing some subsequent act. And it’s reasonable to suppose that the answer to such questions will be relevant for normative ethics regardless of whether the goodness of consequences is the only thing matters (as consequentialists presume) or just one of many things that matter (as non-consequentialists presume)”—Provided by publisher. Identifiers: LCCN 2020018288 (print) | LCCN 2020018289 (ebook) | ISBN 9780190905323 (hardback) | ISBN 9780190905347 (epub) | ISBN 9780190905354 Subjects: LCSH: Consequentialism (Ethics) Classification: LCC BJ1500.C63 O94 2020 (print) | LCC BJ1500.C63 (ebook) | DDC 171/.5—dc23 LC record available at https://lccn.loc.gov/2020018288LC ebook record available at https://lccn.loc.gov/2020018289

1 3 5 7 9 8 6 4 2

Printed by Integrated Books International, United States of America

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Contributors

Edited by Douglas W. Portmore

The Oxford Handbook of Consequentialism

Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy Online Publication Date: Oct 2020

(p. viii) (p. ix) Contributors

Alfred Archer

is Assistant Professor of Philosophy at Tilburg University and a member of the Tilburg Center for Logic, Ethics, and Philosophy of Science. His primary research interests are in moral philosophy and moral psychology, particularly supererogation, the nature and ethics of admiration, and the ethics of fame. He also has research interests in applied ethics, political philosophy, and the philosophy of sport. He is currently working on a project investigating the nature, ethics, and value of admiration, funded by an NWO Veni grant. For up-to-date information about his research, visit http://alfredarcher.weebly.com/.

Calvin C. Baker

is a PhD student in philosophy at Princeton University. His work focuses on ethics, Buddhist philosophy, and global priorities research.

Samantha Brennan

is Dean of the College of Arts and Professor of Philosophy at the University of Guelph. Her research focuses on contemporary normative ethics, including feminist ethics. A recent area of focus for her work is children's rights, parents' rights, and issues of

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

family justice. She's also written and published about micro-inequities, the climate issue in philosophy departments, the moral significance of fashion, and the badness of death.

David O. Brink

is Distinguished Professor of Philosophy at the University of California, San Diego. His research interests are in ethical theory, history of ethics, moral psychology, and jurisprudence. He is the author of *Moral Realism and the Foundations of Ethics* (New York: Cambridge University Press, 1989), *Perfectionism and the Common Good: Themes in the Philosophy of T.H. Green* (Oxford: Clarendon Press, 2003), *Mill's Progressive Principles* (Oxford: Clarendon Press, 2013), and *Fair Opportunity and Responsibility* (Oxford: Clarendon Press, 2021).

Mark Budolfson

is Assistant Professor in Population-Level Bioethics, Philosophy, and Environmental Health Sciences at Rutgers University. He works on issues in philosophy, politics, and economics. Current research includes global ethics and international institutions, population-level bioethics, sustainable development and climate change economics, and reasons for action in collective action situations.

Krister Bykvist

is Professor of Practical Philosophy at Stockholm University, and Research Fellow at the Institute for Futures Studies, Stockholm. His primary interests are in moral philosophy broadly conceived. Most of his research is on topics in normative ethics, including consequentialism, utilitarianism, population ethics, climate ethics, prudence, and well-being. In metaethics, he has done work on noncognitivism, the (p. x) nature of intrinsic goodness, and the normativity of mental states. More recently, he has done work on moral uncertainty. He has coauthored a book on this topic entitled *Moral Uncertainty*, to be published by Oxford University Press (release date February 2020). He has also written a book on utilitarianism, entitled *Utilitarianism: A Guide for the Perplexed* (Continuum, 2010).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Richard Yetter Chappell

is Assistant Professor of Philosophy at the University of Miami. His primary research interests concern the defense and development of consequentialism, effective altruism, and robust normative realism. Chappell blogs at www.philosophyetc.net about these and other philosophical topics. He has published widely in journals, including *Noûs*, *Australasian Journal of Philosophy*, *Philosophical Studies*, and *Philosophical Quarterly*, and was coawarded the Rocky Mountain Ethics Congress 2013 Young Ethicist Prize.

Michael Cholbi

is Professor of Philosophy at the University of Edinburgh. He has published widely in ethical theory, practical ethics, and the philosophy of death and dying. His books include *Suicide: The Philosophical Dimensions* (Broadview, 2011), *Understanding Kant's Ethics* (Cambridge University Press, 2016), and *Grief: A Philosophical Guide* (Princeton University Press, expected 2021). He is the editor of several scholarly collections, including *Immortality and the Philosophy of Death* (Rowman and Littlefield, 2015), *Procreation, Parenthood, and Educational Rights* (Routledge, 2017), *The Future of Work, Technology, and Basic Income* (Routledge, 2019), and *The Movement for Black Lives: Philosophical Perspectives* (Oxford University Press, 2020). He is the founder of the International Association for the Philosophy of Death and Dying and the coeditor of the textbook *Exploring the Philosophy of Death and Dying: Classic and Contemporary Perspectives* (Routledge, 2020). His current research addresses paternalism, assisted dying, and topics related to work and labor.

Yishai Cohen

is Assistant Professor of Philosophy at the University of Southern Maine. His research focuses on agency, ethics, metaphysics, and the philosophy of religion. He is particularly interested in the relationship between libertarian free will and a variety of issues in ethics, including 'Ought' Implies 'Can', the Principle of Alternate Possibilities, and the actualism/possibilism debate.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Dale Dorsey

is Dean’s Professor and Chair of the Department of Philosophy at the University of Kansas. He generally works in normative ethics, at the intersection of the personal good, morality, and practical rationality. He has also worked on metaethics and has written essays on the moral philosophy of David Hume, Francis Hutcheson, and John Stuart Mill.

Julia Driver

is Professor of Philosophy at the University of Texas at Austin and Professorial Fellow at the Centre for Ethics, Philosophy, and Public Affairs at St. Andrews. Her research is primarily focused on normative ethics, metaethics, and moral psychology. She is the author of several books, the most recent being *Consequentialism* (Routledge, 2012).

(p. xi) **Hilary Greaves**

is Professor of Philosophy and Director of the Global Priorities Institute at the University of Oxford. Her main research interests concern issues in moral philosophy, decision theory, and economics, with a special focus on issues that arise in the course of considering how an altruistic actor might most cost-effectively do good. Her published work includes articles on moral uncertainty, population ethics, discounting, and theories of well-being and of interpersonal aggregation.

Matthew Hammerton

is Assistant Professor of Philosophy at Singapore Management University. He has published several articles on the structure of moral theories such as consequentialism, deontology, and virtue ethics.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Contributors

Caspar Hare

is Professor of Philosophy at the Massachusetts Institute of Technology. He writes about ethics, practical rationality, metaphysics, and about the connections between them. He is the author of two books: *On Myself, and Other, Less Important Subjects* (Princeton University Press, 2009) and *The Limits of Kindness* (Oxford University Press, 2013).

Brad Hooker

is Emeritus Professor at University of Reading and a Senior Research Fellow at Uehiro Centre for Practical Ethics at University of Oxford. He has published on a wide array of topics in ethics but is best known as the author of *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*.

Paul Hurley

is the Sexton Professor of Philosophy at Claremont McKenna College, The Claremont Colleges. His research focuses primarily upon ethics, particularly the debate between consequentialists and their critics, but he has also published in metaethics, action theory, and the history of ethics. He is the author of *Beyond Consequentalism* (Oxford University Press, 2009) and of over two dozen articles. His current project is to demonstrate that the central arguments for consequentialism are grounded below ethics, in outcome-centered theories of actions and attitudes, and to challenge the case for consequentialism at this deeper level.

Frank Jackson

is Emeritus Professor at The Australian National University. He works in the philosophy of mind, ethics, and the philosophy of language. His books include *Conditionals* (Blackwell, 1987), *From Metaphysics to Ethics* (Oxford, 1998), and *Language, Names, and Information* (Wiley-Blackwell, 2010).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Diane Jeske

is Professor of Philosophy at the University of Iowa, where she has taught since 1992. Her work has focused on the nature and significance of intimate relationships and how that significance ought to be reflected in moral theory. She is the author of *Rationality and Moral Theory: How Intimacy Generates Reasons* (Routledge, 2008), *The Evil Within: Why We Need Moral Philosophy* (Oxford University Press, 2018), and *Friendship and Social Media: A Philosophical Exploration* (Routledge, 2019).

Tyler M. John

is a PhD student in philosophy at Rutgers University–New Brunswick. His main areas of research are distributive ethics, political philosophy of the long-term future, and animal moral, legal, and political philosophy. He is a coauthor of *Chimpanzee Rights: The Philosophers' Brief* (2018) and of articles appearing in *Ethics* and *Economics and Philosophy*. *(p. xii)*

Victor Kumar

is Assistant Professor at Boston University. He works mainly at the intersection of ethics and cognitive science. His published work can be found in *Ethics*, *Noûs*, and *Philosophers' Imprint*. In recent years he has written about moral learning, moral luck, and moral disgust. He is currently writing a book with Richmond Campbell about moral evolution and moral progress.

Holly Lawford-Smith

is a Senior Lecturer in Political Philosophy at the University of Melbourne. She works on topics across moral and political philosophy, applied ethics, and social ontology, including climate ethics, corporate responsibility, collective agency, and radical femi-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

nism. Her first book *Not in Their Name: Are Citizens Culpable for Their States' Actions?* came out with Oxford University Press in 2019.

Katarzyna de Lazari-Radek

is Assistant Professor at the Faculty of Philosophy, University of Łódz, Poland. She is a hedonistic utilitarian. Her main research interest focuses on the philosophy of Henry Sidgwick and Derek Parfit, as well as the concept of well-being and pleasure. Together with Peter Singer she wrote two books: *The Point of View of the Universe* (Oxford University Press, 2014) and *Utilitarianism—A Very Short Introduction* (Oxford University Press, 2017). Apart from academic work, she is keen to convey philosophical ideas to a wider audience, giving lectures and writing for popular magazines on how to live a good life.

Alida Liberman

is Assistant Professor of Philosophy at Southern Methodist University. Her main research interests are in theoretical and applied ethics, and she is particularly interested in how our attitudes and commitments affect what it makes sense for us to do. She has published papers about promises, vows, endorsements, and resolutions, as well as about a variety of topics in bioethics, and is currently working on a project about the ways in which it is wrong to make it harder for others to fulfill their obligations.

Judith Lichtenberg

is Professor Emerita of Philosophy at Georgetown University. Her primary fields of interest are international and domestic justice, moral psychology, nationalism, war, and higher education. Her book *Distant Strangers: Ethics, Psychology, and Global Poverty* was published by Cambridge University Press in 2014. With Robert Fullinwider, she coauthored *Leveling the Playing Field: Justice, Politics, and College Admissions* (2004); she is the editor of *Democracy and the Mass Media* (1990). For the last several years she has been teaching philosophy at Jessup Correctional Institution in Maryland and at the D.C. Jail in Washington.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Contributors

Barry Maguire

is Assistant Professor of Philosophy at Stanford University. Previously, he taught Politics, Philosophy, and Economics at UNC Chapel Hill, and held a Bersoff Fellowship at NYU. He works on issues at the intersection of normative theory, normative ethics, political ethics, and the ethics of economics.

Elinor Mason

is Professor of Philosophy at the University of California, Santa Barbara. She works on ethics, moral responsibility, and feminist philosophy. She is the author of *Ways To Be Blameworthy: Rightness, Wrongness, and Responsibility* (Oxford University Press, 2019).

(p. xiii) **Joseph Mendola**

is Professor of Philosophy at the University of Nebraska—Lincoln. His research interests include ethics, metaphysics, and philosophy of mind. He is the author of four books: *Human Thought* (Kluwer, 1997), *Goodness and Justice* (Cambridge University Press, 2006), *Anti-Externalism* (Oxford University Press, 2008), and *Human Interests* (Oxford University Press, 2014).

Shyam Nair

is Assistant Professor of Philosophy at Arizona State University. He is primarily interested in issues in ethics, epistemology, and philosophical logic. His research focuses on formal and philosophical questions at the intersection of these fields concerning how best to model what we ought to do, what we ought to believe, and how we ought to reason.

Alastair Norcross

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

is Professor of Philosophy at the University of Colorado Boulder, where he has taught since 2007. Prior to that, he taught at Southern Methodist University and Rice University (before being allowed out of Texas for good behavior). He works both on ethical theory and on issues in applied ethics. In ethical theory he has published extensively on consequentialism, in particular defending a scalar version of the theory. His book *Morality by Degrees: Reasons without Demands* (Oxford University Press, 2020) articulates and defends the scalar approach. In applied ethics he has published many articles criticizing the common practices of raising animals for food and using them in experimentation, including the widely reprinted “Puppies, Pigs, and People: Eating Meat and Marginal Cases” (*Philosophical Perspectives*, 2004). He also runs marathons, with somewhat less success than Eliud Kipchoge, and writes, directs, and acts in the theater, with somewhat less success than Kenneth Branagh.

Douglas W. Portmore

is Professor of Philosophy at Arizona State University. His research focuses mainly on morality, rationality, and the interconnections between the two, but he has also written on blame, well-being, moral worth, posthumous harm, moral responsibility, and the nonidentity problem. He is the author of two books: *Commonsense Consequentialism: Wherein Morality Meets Rationality* (Oxford University Press, 2011) and *Opting for the Best: Oughts and Options* (Oxford University Press, 2019).

Melinda A. Roberts

is Professor of Philosophy at the College of New Jersey and recently completed a Laurance S. Rockefeller faculty fellowship at the Princeton University Center for Human Values. Both a philosopher and a lawyer, she is the author of *Child Versus Childmaker*, *Abortion and the Moral Significance of Merely Possible Persons* and a number of articles in the areas of population ethics (including the repugnant conclusion and the nonidentity problem), procreative ethics (including wrongful life and reproductive technologies), and climate ethics. She continues to have an interest in developing a person-based form of consequentialism that functions well for both the evaluation of choices and of outcomes.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Jeff Sebo

is Clinical Associate Professor of Environmental Studies, Affiliated Professor of Bioethics, Medical Ethics, and Philosophy, and Director of the Animal Studies M.A. Program at New York University. He works primarily on bioethics, animal ethics, and environmental ethics. His coauthored books *Chimpanzee Rights* and *Food, Animals, (p. xiv) and the Environment* are currently available from Routledge, and his book *Why Animals Matter for Climate Change* is currently in contract with Oxford University Press. Jeff is also on the Board of Directors at Animal Charity Evaluators, the Board of Directors at Minding Animals International, and the Executive Committee at the Animals & Society Institute.

Holly M. Smith

is Distinguished Professor Emerita of Philosophy at Rutgers University and Distinguished Research Associate at The University of California, Berkeley. She has also held appointments at Tufts University, the University of Pittsburgh, the University of Michigan, the University of Illinois-Chicago, and the University of Arizona. Her publications principally focus on topics in normative ethics, moral decision making, the theory of moral responsibility, and biomedical ethics. In *Making Morality Work* (Oxford University Press, 2018), she explores how moral theories should accommodate the errors, ignorance, and misunderstandings that impede us as moral decision makers. Her current projects propose new strategies for weighing the stringency of deontological duties, and for identifying and evaluating an agent's alternatives in the context of normative theories.

David Sobel

is Guttag Professor of Ethics and Political Philosophy at Syracuse University. He is the author of *From Valuing to Value* (Oxford University Press, 2017), founding coeditor of the *Oxford Studies in Political Philosophy* series, and coeditor of the blog *PEA Soup*. His primary research project focuses on the question of what makes things valuable. He is especially interested in whether, and to what extent, it is our attitudes toward things that make them valuable for us.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Dean Spears

is an economic demographer and development economist. His research areas include the health, growth, and survival of children in developing countries and population dimensions of social well-being. Dean is Assistant Professor of Economics at the University of Texas at Austin, is a visiting economist at the Economics and Planning Unit of the Indian Statistical Institute in Delhi, is a founding Executive Director of r.i.c.e. (a nonprofit that works for children's health in India), and is an affiliate of IZA, of the Institute for Futures Studies, and of the Climate Futures Initiative at Princeton University. With Diane Coffey, he is a coauthor of the book *Where India Goes: Abandoned Toilets, Stunted Development, and the Costs of Caste* (2017); he is the author of *Air: Pollution, Climate Change, and India's Choice between Policy and Pretence* (2019). His research is supported by an NIH Population Scientist career grant.

Travis Timmerman

is Assistant Professor of Philosophy at Seton Hall University. He specializes in normative ethics, applied ethics, and the philosophy of death. In normative ethics, he primarily works on the actualism/possibilism debate, having recently coauthored the *Stanford Encyclopedia of Philosophy* (2019) entry on the topic as well as "How To Be an Actualist and Blame People" in *Oxford Studies in Agency and Responsibility* (2019). In the death literature, he focuses on axiological questions about death's badness. Recent publications include "A Dilemma for Epicureanism" in *Philosophical Studies* (2019) and "Avoiding the Asymmetry Problem" in *Ratio* (2018). In applied ethics, he works on issues related to global poverty and questions about (p. xv) the ethics of historical monuments. Publications in applied ethics include "Sometimes There Is Nothing Wrong with Letting a Child Drown" in *Analysis* (2015) and "A Case for Removing Confederate Monuments" in Oxford University Press's *Ethics Left and Right* (2020).

William Tuckwell

is a PhD candidate in Philosophy at The University of Melbourne. His research focuses mainly on social and political philosophy, epistemology, and the interconnections between the two.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Contributors

Christopher Woodard

is Professor of Moral and Political Philosophy at the University of Nottingham, UK, and President of the British Society for Ethical Theory. His research focuses on consequentialism (especially collective forms of consequentialism), well-being, and normative reasons for action. He has also written on other topics in moral and political philosophy, including egalitarianism, meaning in life, and the actualism-possibilism debate. He is the author of two books: *Reasons, Patterns, and Cooperation* (Routledge, 2008) and *Taking Utilitarianism Seriously* (Oxford University Press, 2019). (p. xvi)

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Consequences

Dale Dorsey

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.1

Abstract and Keywords

The consequences of our actions seem to matter. But what is the nature of the consequence relation that a particular act bears to, well, its consequences? This essay considers a number of traditional approaches to understanding the consequence relation. While many traditional approaches treat the consequence relation as built upon a *causal* relation, I hold that there are good reasons to doubt that the consequence relation should be understood in terms of causal relations, even if supplemented with the identity relation. Instead, I argue for a contrastive approach that, while not entirely free of problems, does a better job than standard accounts at capturing the relationship between an act and its consequences.

Keywords: consequences, cause, G. E. Moore, action, omission

A pivotal scene in the film *The Blues Brothers* has Jake and Elwood Blues arrive at a Chicago restaurant in the hopes of recruiting the cook (Matt “Guitar” Murphy) and dishwasher (“Blue” Lou Marini) to rejoin their rhythm and blues revue. Set against this notion is the proprietor of the establishment, played by Aretha Franklin. As Murphy is set to leave, Franklin[1] turns to him and says: “You better think about what you’re sayin’! You better think about the consequences of your actions!”

This seems like advice worth heeding. In deciding what to do, good practical reasoners are often supposed to take the consequences of their actions very seriously. And while it may be that good decision-making is not entirely taken up with consideration of the consequences of the various lines of conduct open to an agent, rational thought about how to act is certainly, at least to some degree, taken up with the difference our act makes to the world at large.

Indeed, the consequences of actions seem relevant to the evaluation of such actions in virtually all (if not actually all) domains of evaluation. Surely the consequences (and their evaluative valence) will at least help to determine whether an action was prudent, or aesthetically justified, or morally acceptable. In this essay, I’m interested in exploring what, precisely, *are* the consequences of an action. What, put more precisely, is the content of

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

the *relation* between a particular state of affairs S and an action ϕ such that S is rightly described as a consequence of ϕ?

The plan of the paper runs like this: after briefly discussing the significance of consequences in various domains of inquiry, I then go on to discuss a number of extant accounts of the consequence relation, including most importantly what I call the "traditional" account of an act's consequences. After rejecting the traditional approach, I defend a "contrastive" view, according to which the consequences of an act can only be identified relative to other acts and are determined in such a contrast by the difference in (p. 94) *extrinsic properties* between the act and its contrast (given an assumption of a relevant set of background conditions). I defend this view from objections and argue that this approach can capture what we take ourselves to be normatively interested in when we are interested in the consequences of an act.

1. The Significance of Consequences

The consequences of an action seem, well, important. Take morality. The following principle seems darn-near incontrovertible:

> *Minimal Objective Consequentialism*: the fact that an act ϕ's consequences are better than another act ψ's consequences is an objective moral reason in favor of ϕ rather than ψ.

For *Minimal Objective Consequentialism*, the quality of the consequences of a given act relative to other acts constitutes an objective moral reason to favor the former rather than the latter. Two notes concerning the *minimalism* of *Minimal Objective Consequentialism*. First, it is important that *Minimal Objective Consequentialism* concerns *objective* moral reasons. By "objective," here, I mean simply a perspective of moral evaluation that is abstracted from the epistemic circumstances of the agent in question. They are moral reasons, put bluntly, that we would expect the agent to conform to if only the agent *knew all the facts*. Of course, *Minimal Objective Consequentialism* would be far more than minimal if it were not limited in this way. It is perfectly sensible to hold that at the very least subjective moral reasons will abstract from the actual consequences of an action to focus instead on consequences that were or could have been foreseen or predicted, given the agent's evidence or available evidence. Second, note that *Minimal Objective Consequentialism* does not say that there are no other objective moral reasons, taken from non-consequence-based facts about the act. *Minimal Objective Consequentialism* does not rule out, for instance, motives, conformity to rules, respect for persons, or any other such factor as mattering to the objective moral quality of an action. It merely says that if two actions differ in the goodness of their consequences, this is but *one* reason to favor the action with better consequences.

But notice that even beyond morality, the consequences of actions clearly seem to have significance in determining how to act. Consider, for instance, prudence. While there are a number of controversies that surround prudential evaluation of action,[2] one settled mat-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ter seems to be that prudence has a consequentialist character: actions are prudentially favored to the extent that their consequences are better for the agent. Here actions are evaluated given the quality of their consequences *for the agent*.

(p. 95) Furthermore, prudence is not the only domain that seems to evaluate action primarily on grounds of an act's consequences. Take also *aesthetics*. Of two possible acts, it seems right to hold that there is reason *of aesthetics* to choose that act that will in fact increase the *beauty* of the resulting states of affairs. That, in other words, the consequences of an act as assessed by their beauty is relevant to the aesthetic evaluation of such acts. Given this, it would seem sensible to ascertain what the nature of the consequence relation is, not least of which because it may help us to determine whether such plausible claims hold up on reflection. Put another way, while we generally regard consequences as important, we cannot *really* know whether, for example, *Minimal Objective Consequentialism* stands up to considered judgment until we know just what, in fact, a consequence of an action really is.

2. Theories of the Consequence Relation

When "Guitar" Murphy leaves the restaurant with the Blues Brothers, surely one of the consequences of this action is that Aretha Franklin is made angry, customers are not served, and so on. But in virtue of what are these states consequences of Murphy's departure?

Note that I don't mean to ask the question: what *sort of thing* can *qualify* as a consequence, that is, can consequences be states of affairs, events, processes, actions, and so forth? For the remainder I'm going to abstract from that question (I'll assume that only states of affairs are consequences here, though this makes no never mind here). Instead, I'm interested in the nature of the consequence *relation*: what relation must a particular, for example, state, event, and so forth bear to a particular act to be properly assigned as a consequence of that act?

2.1. The Moralized Approach

In the face of the importance of consequences in evaluating the quality of acts, it may seem like the obvious answer is to link the consequence relation to that very significance. What I shall call the "moralized approach" (but which also could be easily translated into the "prudentialized" or "aestheticised," or whatever, approach) holds that the consequence relation between an act and a state of affairs should be determined in moralized terms—terms, that is, that are typically accepted by paradigmatically consequentialist theories. To illustrate one example, D. D. Raphael writes: "This, then, is my first 'conclusion': the consequences of an action are those results for which the agent is held to be (p. 96) responsible, those results of which his action is 'the cause'."[3] H. L. A. Hart and A. M. Honoré write:

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> The Utilitarian assertion that the rightness of an action depends on its consequences is not the same as the assertion that it depends on all those later occurrences which would not have happened had the action not been done, to which indeed "no limit can be set" …[W]henever we are concerned with such connexions, whether for the purpose of explaining a puzzling occurrence, assessing responsibility, or giving an intelligible historical narrative, we employ a set of concepts restricting in various ways what counts as a consequence …[T]he voluntary intervention of a second person very often constitutes the limit. If a guest sits down at a table laid with knife and fork and plunges the knife into his hostess's breast, her death is not in any context thought of as caused by, or the effect or result of the waiter's action in laying the table; nor would it be linked with this action as its consequence for any of the purposes, explanatory or attributive, for which we employ causal notions.[4]

Here Hart and Honoré suggest that the voluntary intervention of a second person constitutes the limit of the consequences of a particular action, in part (or so one assumes) because we do not wish to hold the waiter in any way responsible for the murder of the hostess.

However, moralized accounts have a number of ill results. Take, for instance, Raphael's view. Perhaps most obviously this view would seem to make a sensible claim, viz., "we don't hold people responsible for all the consequences of their actions" *conceptually* false. But surely even if this claim is false, it is not *conceptually* false. Indeed, it's not even clear to me that it is false. If, for instance, we do not hold people responsible for consequences they might not have known about or foreseen (to put this another way, we typically hold people responsible only for violations of *subjective* moral reasons), it seems entirely false that all the consequences of an action are those that must be linked to that for which we would hold an agent responsible. Hart and Honoré's view—admittedly underspecified—is even more problematic in this regard. To take a simple example,[5] consider the action of Adolf Hitler's great-great-great-great grandmother in deciding to have Adolf Hitler's great-great-great grandfather. The Beer Hall Putsch could not have been foreseen by this person, nor was it intended. But is it a consequence? Of course. Furthermore, their own example seems implausible on reflection. While we may not wish to hold the waiter responsible for the murder of the hostess, it's clear that this murder was, in fact, a consequence of his setting the table in such-and-such a way (if, or so we are given to assume, the murder would not otherwise have occurred).

(p. 97) 2.2. A Traditional Approach

However, there is a proud tradition of accounting for the consequences of an action that abstracts from moralized concerns. G. E. Moore puts it in the following terms:

> One natural way, and perhaps the most natural way, of understanding the expression "the total consequences of the action, A" is one in which among the consequences of A nothing is included but what is the case *subsequently* to the occur-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> rence of A, so that the "total consequences of A" means everything which is the case *subsequently* to A's occurrence, which is also such that it would not have been the case if A had not occurred.[6]

A note on Moore's proposal should be made here. First, Moore is discussing the *total* consequences of a particular action—all of the states of affairs that *are* consequences. But this can easily be abstracted away. Moore's account of the consequence relation in this passage seems to run as follows: S is a consequence of ϕ if S happened subsequent to ϕ and would not have happened had ϕ not occurred.

Perspicuous though it is, we should immediately be suspicious of Moore's account. To begin, to adopt a temporal constraint on the nature of consequences would be a mistake. Or, at the very least, a mistake in understanding the conceptual nature of the consequence relation. Imagine, for instance, that we countenance the possibility of backward causation. Suppose some action ϕ at time t would *cause* some thing (some state, say) S at time t_{-1}. Surely the right proposal here is that S is properly a consequence of ϕ. Of course, whether backward causation is possible seems, to my mind, a little unlikely. But this point shouldn't make any difference to the very structure of the consequence relation itself.

But why is this? Why should we allow that a past state could be a consequence given the possibility of backward causation? A plausible answer lies in Moore's insistence that consequences are "such that [they] would not have been the case if A had not occurred." To put this another way, consequences of a particular action are intended to capture the *difference* that action makes to the world—that which different about the world *given* the action. How did A's ϕ-ing change, or not change, the world? Moore's insistence on a temporal constraint is, of course, quite natural *given* the assumption that any changes to the world given a particular act will happen after the act. But insofar as we wish a more general account of the consequence relation that abstracts from debates on whether there can be backward causation, we should instead focus on what it means in a more general sense for the act to *make a difference*.[7]

But what does it mean to "make a difference" in this way? Perhaps the most obvious candidate given our discussion so far is for the action to *cause* some state of affairs or (p. 98) other. After all, in assigning consequences to Murphy's decision to rejoin the Blues Brothers, we naturally look to what this action *caused*.[8] Of course, the nature of the causal relation is philosophically contested. But it may be that a proper account of the metaphysics of consequences can simply ride "above the fray" on this question and simply stipulate that the proper relation here just is whatever the proper causal relation turns out to be.

However, there is a further method by which the act could make a difference to the world. And that is, for it to, well, *be*. Many hold that the action itself should be understood as a consequence of that act, insofar as it is certainly one feature of the world for which the act itself made a difference.[9] But this criterion cannot be assimilated into the causal condition. Acts, after all, are not their own causes.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Given its pedigree, I'm going to call the view according to which an act's consequences are the *causal effects* of the act and the *act itself* the *traditional view*. In what follows, I suggest a number of potential worries about the traditional approach.

3. Concerns about the Traditional Approach

I think there are sensible worries to be had about both ends of the traditional account. Begin with the *causal* bit.

3.1. Why a Causal Relation?

The traditional approach to identifying the consequences of an action holds that some state of affairs (suppose) S is a consequence of an action ϕ if and only if (a) S is ϕ or (b) ϕ caused S. Ignore, for present purposes, the identity claim here. Concentrating strictly on the right-hand side of the traditional disjunction, there may be reasons to be concerned about whether (a) all (what I shall call) "nonidentical" consequences (i.e., consequences that are not simply the act itself) must be caused by the act[10] and (b) whether even all causal effects of an act are, in fact, consequences.

With regard to the first point, take Frances Kamm's notion of a *noncausal flip side*. Referring to the infamous Trolley Case, she writes:

(p. 99)

> In the Trolley Case, intuitively we think that we may redirect the trolley that causes the death of the one. In the context where only the trolley threatens the five, the redirection of the trolley threat away from them—by which I mean the moving of the trolley itself away, is a means of saving them. It is the same event as their becoming free of threats, and this is the same event as their becoming saved. Put another way, in the context of the original Trolley Case, the five being free from threats is constituted by the trolley being away from them. Hence, there is a *noncausal* relation between the trolley turning away and the five being saved. Because this is true, I will say that the five being saved, which is the greater good when continuing life is good for the five, is the noncausal flip side of the redirection (in my sense of the moving away) of the trolley.[11]

For Kamm, the saving of the five was not *caused* by the turning of the trolley, because the turning of the trolley was, in fact, the same event as their "becoming saved," though it was clearly a means to that event. But surely we would like to say that the fact that five were saved *is a consequence* of turning the trolley. If this is right, it would appear that there are states that are nonidentical with the act, not *caused* by the act, but that are nevertheless consequences.

Of course, one might be tempted to quibble with Kamm's suggestion that the saving of the five is constituted *by* the turning of the trolley, or that the saving of the five is not caused by the turning of the trolley. But I propose to leave the possibility of noncausal flip

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

sides and to instead focus on the possibility of noncausal, nonidentical consequences of *omission*. To illustrate my concern, consider:

> *Gerald*: Gerald is considering whether to share his ice cream with his best friend or to eat it all himself. Beset with indecision, he does nothing. His ice cream melts, and neither Gerald nor his friend gets any ice cream.

Here's an initial reaction to this case. Gerald's failure to do anything with the ice cream cone didn't *cause* the ice cream to melt. How could it? He didn't do anything. But *that* the ice cream melted and neither Gerald nor his friend got any ice cream is certainly a *consequence* of Gerald's failure to do something with it (either eat it or share it). And so for a nonidentical state of affairs to be a consequence of a given act (i.e., doing nothing) does not require that the act serve as cause of that state of affairs.

Response: why should it be that Gerald's failure to do anything didn't cause the ice cream to melt? Indeed, there is a growing literature dedicated to understanding the nature of *omissions* as causes. The classic example runs as follows. X might ask Y to water X's plants while X is on vacation. Y forgets, and the plants die.[12] It seems plausible to say, under such conditions, that Y's failure to water the plants caused them to die. Y, in this case, didn't do anything. But in not doing anything, Y caused the plants to die. And hence we may be tempted to say that even if Gerald didn't do anything at all, his failure to (p. 100) do something is causally responsible for the melting of the ice cream. Hence there is no problem in understanding the consequences of Gerald's failure as a *causal* relation.

But there is a serious complication here. The literature on omissions as causes faces what is known as the *selection problem*. After all, there are *many* omissions that seem to stand in a very similar relation to the melting of Gerald's ice cream as Gerald's indecision. For instance, no bystanders stole the nearly melted ice cream cone and rushed it to Gerald's friend, or decided to consume it themselves. But it would be strange to say that the failure of a bystander to steal Gerald's rapidly melting ice cream cone *caused it to melt*. On this point, Sara Bernstein writes:

> the counterfactual account of causation, upon which *c* is a cause of *e* if *e* counterfactually depends on *c*, admits far more omissions as causes than are intuitively so. Unlike oomph-y theories of causation, omissions easily fit into counterfactuals of the form
>
> ¬A ->¬C
>
> such as
>
> If the technician hadn't failed to perform the safety check, the plane wouldn't have crashed.
>
> But other omissions, such as

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> If Barack Obama hadn't failed to perform the safety check, the plane wouldn't have crashed.
>
> also generate counterfactual dependence between putatively irrelevant omissions (such as Barack Obama's failing to perform the safety check) and the effect. Such omissions come out as causes on the counterfactual view, even when they are not, intuitively, causes … Admitting *any* omissions seems to admit *all* omissions—even the nonsalient ones—as causes.[13]

Bernstein puts this as a problem specifically for counterfactual dependence theories of causation. But this is irrelevant for present purposes. It seems intuitively correct that Barack Obama's failure to perform the safety check did not cause the plane to crash just as, intuitively, it seems right to say that the bystander failing to steal Gerald's ice cream cone didn't cause it to melt. Rather (if we seek to speak of omissions as causes at all), it was the omission of the safety technician that caused the plane to crash, and Gerald's failure to do anything that caused the ice cream cone to melt. The selection problem, or at least this version of the selection problem, applies to any account of the way in which omissions can be causes. We must divide up the *right* omissions from the *wrong* ones—the *really causal* ones as opposed to the omissions that don't have the right relation to the caused event. To put this in a slightly different way, the *true theory of causation* will allow that Gerald's failure to do anything caused the ice cream to melt, but it will *not* allow that the bystander's failure to steal the ice cream caused the ice cream to melt.

(p. 101) Note that there are lots of attempts to do this.[14] I'm not going to comment on whether any of them are successful. But the point I wish to make is that, while it is clearly unintuitive to say that the bystander's failure to steal the ice cream cone *caused* it to melt, it is nevertheless quite intuitive to say that the melting of the ice cream cone was a *consequence* of the bystander's failure to steal the ice cream cone. Indeed, surely one consequence of the fact that Barack Obama failed to perform the relevant safety check was the plane's crash. Now, of course, the normative significance of such consequences are another matter. But clearly there is a difference in the world made by Barack Obama's omissions, as well as the bystander to Gerald's indecision. And hence the consequence relation seems to extend beyond relations that can be legitimately understood as *causal*.

So even if we abstract from the act itself, it's not clear that the structure of the consequence relation is causal. But, furthermore, it's not clear that, even if we have established a clear causal relation between an act and a state of affairs, the latter need be a consequence of the former. To see this, consider:

> *Firing Squad*: A prisoner stands before a firing squad of six soldiers. They all fire with deadly accuracy, but the fourth soldier's bullet arrives first.

Plausibly, the fourth soldier's firing of the gun *caused* the death of the prisoner. But was the death of the prisoner (more precisely, the state of affairs in which the prisoner is dead) a *consequence* of the fourth solder's act? Plausibly, no. This is because the fourth

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

soldier's act did not itself stand between the survival of the prisoner and the death of the prisoner. Had the fourth soldier not fired, the prisoner would have died by one of the other bullets.

Now, that's not to say that the fourth soldier's act had no consequences with regard to the death of the prisoner. Plausibly, it had the consequence of the prisoner's death *at such-and-such a spatio-temporal location*, let's say. But though it *caused* the prisoner's *death by firing squad*, the prisoner's death by firing squad was not a consequence—it made no difference to whether the prisoner died by firing squad or did not.

Thus, or so it seems to me, consequences need not be causally related to a particular act, and not all things related *via* causation to the act are, in fact, consequences. It would appear, then, that the causal relation is only an accidental feature of some instances of the consequence relation.

3.2. Why the Act Itself?

When Matt "Guitar" Murphy leaves his job as a cook, this action surely has a number of consequences. He rejoins the Blues Brothers. He angers Aretha Franklin. But is it right to say that one consequence of Murphy's leaving his job is that he *leaves his job*?

(p. 102) Frankly, this sounds a little strange to me. Why should it be that the act itself is included as part of the *consequences* of the act? Put more precisely, it would seem that the consequence relation is *nonreflexive*. If, in other words, ϕ is the consequence of ψ, this implies that ψ is not the consequence of ϕ. Surely this is borne out by traditional sorts of examples. A consequence of "Guitar" Murphy's leaving his job was that the Blues Brothers saved an impoverished orphanage. But not *vice versa*. A consequence of the trolley being turned is that the five were saved. But not *vice versa*.[15] But, of course, this would fail were the act itself its own consequence. Here's another way to put the argument of this paragraph: paradigmatic examples of the consequence relation go one way only. If this is correct, it would seem quite strange to believe that *in the case of the act itself*, the consequence relation goes two ways. After all, paradigmatic examples establish this relation as nonreflexive.

Jonathan Bennett, however, challenges this. According to Bennett, those who would accept a key distinction between the act itself and the consequences of the act would hold that the act itself is composed of, perhaps obviously, its intrinsic properties, but that the consequences are composed only of its relational properties, a particular subset, that is, of its relational properties. But:

> It seems [odd] when one recalls how impoverished are the strictly intrinsic facts about how people behave. Nearly everything we say about behaviour attributes relational properties to it, and that includes everything with moral significance. An intrinsic account of how Agent behaved would give only the geometrical or balletic qualities of his movements. Such facts might matter if we held it to be funda-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> mentally wrong to make (say) circular movements with one's left hand; but nobody accepts such a morality.[16]

Bennett, I think, clouds the issue somewhat (at least for our purposes) by referring specifically to what we might find morally significant, that is, the act itself or its consequences. But there's a point here that's worth taking seriously. When we say that Matt "Guitar" Murphy left his job, we don't *really* seem to be describing the intrinsic properties of his act. His act consists simply in his moving his arm in a certain way (i.e., to throw his apron on the floor) and then using his legs to walk in a certain direction. But if this is right, then what we *describe* of the action (he "leaves his job") *does* seem to be a *consequence* (in Kamm's language, a noncasual flip side): he leaves his job *by* moving his arms and legs thus and so. Bottom line: when it comes to the actions that really matter or that we're interested in discussing, it's hard to separate the consequences from the act itself in a principled way.

(p. 103) However, I think this argument is too quick. Note that we often talk about acts under various *descriptions*.[17] One can *describe* Matt "Guitar" Murphy's action as moving his arms and legs thus and so, but one can also describe his action, with its attendant intrinsic properties, as leaving his job. Now, different accounts of act individuation will tell us whether or not these descriptions point to different acts or point to the same act. But this makes no difference here. The key is that *whatever* description we adopt for a given action, we can talk about the act itself *under that description* and the consequences of the act *under that description*. One of the consequences of Murphy's moving his arms and legs thus and so is that he leaves his job. One of the consequences of Murphy's *leaving his job* is that … , and so on. Thus I think we have ample reason to resist the suggestion that the consequence relation should be reflexive, and hence good reason to reject the thought that the act itself could be a consequence of the act.

Given all this, we should reject the traditional account. First, the act itself is not plausibly a consequence of the act. Second, the causal relation, while perhaps a common way to tie the act to its consequences, is neither necessary nor sufficient to establish a consequence relation. But this leaves open the nature of the consequence relation: how are, for instance, the plane crash and Barack Obama's omission related, such that the former can be said to be a consequence of the latter?

4. Reconsidering the Difference Made

Recall the general thought that the consequences of an act are the differences the act makes to the world (that is, independently of its own existence). Surely there is something right about this proposal, but the *traditional* interpretation of this platitude seems (as already argued) inadequate to the consequence relation.

In this section, I'm going to offer an alternative interpretation of this thought, and hence an alternative account of the consequence relation. Now, this proposal is purposefully underspecified: there are a number of ways this view could be refined or modified in re-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

sponse to intuitive data. But to begin articulating my account, let us consider whether the fact that the Blues Brothers played a concert at the Palace Hotel Ballroom is a consequence of Matt "Guitar" Murphy leaving his job. How do we determine this? If the Blues Brothers were going to play a concert at the Palace Hotel Ballroom whether "Guitar" Murphy left his job or not, it seems clear that this state of affairs is not such a consequence. So to accurately claim that this concert is a consequence of his leaving the job, that this, in other words, is part of the difference to the world made by leaving his job, we must *compare* what happens when he leaves his job, with some other state of affairs—we compare it with what otherwise would have happened.

(p. 104) But what does this mean? What *otherwise would have happened* is, obviously, *something*. And it is the content of this something that seems crucial in determining the consequences of any given act, as we have just seen (i.e., whether the "otherwise" includes the Palace Hotel Ballroom gig). But what *is* this "otherwise"? What are we considering as a contrast to "Guitar" Murphy leaving his job? One obvious possibility is the state of affairs in which he does nothing. But this is problematic. For instance, we could simply stipulate that were "Guitar" Murphy not to leave his job, he would contact another guitarist he knows and persuade this other guitarist to play with the Blues Brothers at the Palace Hotel Ballroom. In this case, the Palace Hotel Ballroom gig would occur whether or not Murphy left his job, and hence it is not plausibly a consequence of leaving. But, if we compare "Guitar" Murphy's leaving his job with doing *nothing* (just, say, being frozen in suspended animation for some period of time), there *is* a difference made in the world by his leaving, viz., the Palace Hotel Ballroom gig. And hence this view seems to overpredict the consequences of Murphy's action.

I'm going to cut to the chase here. I think the attempt to find a *privileged* contrast class to fill out the conceptual structure of the consequence relation is a futile line of inquiry. Two reasons. First, while it's clear there needs to be a contrast in understanding the nature of the consequence relation, it's not clear that, to understand that consequence relation, we must ex ante determine the most relevant contrast. To see this, note that we could certainly say that the Blues Brothers' gig at the Palace Hotel Ballroom was a consequence of Murphy's leaving his job *relative to*, say, staying in his job and doing nothing else. But we could also say, with perfect uprightness, that the gig was not a consequence of his leaving *relative to* his working assiduously to find a replacement guitar player. Your feeling a tremendous amount of pain is a consequence of my kicking you in the right shin relative to my giving you a pat on the back, but it is not a consequence of my kicking you in the right shin relative to my kicking you in the left shin. And while there may be a "privileged" understanding of "what otherwise would have happened" in certain contexts (i.e., contrasts we're most interested in in particular domains of inquiry or discourse), this doesn't appear to be the case when it comes to understanding the consequence relation *itself*.

Second, note that I have so far been concentrating on the consequences of *acts*. But acts aren't the only things that have consequences. To see this, consider the "Consequence Argument" against compatibilism, stated originally by Peter van Inwagen:

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> If determinism is true, then our acts are consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things are not up to us.[18]

Here it would appear perfectly cogent to talk about the consequences of particular *laws of nature*, together with certain initial conditions. But what *on Earth* could possibly be (p. 105) the "privileged" contrast class? Presumably, we require comparison with some *other* set of the laws of nature. But which? Any? All?

These questions seem intractable because, or so I submit, they are. Rather than insisting that the conceptual structure of the consequence relation determines one privileged contrast class (rather than leaving that to, say, contextual factors), we should instead say that the conceptual structure of the consequence relation requires *some* act alternative, but that we need not fix on any privileged alternative in determining the consequences of said act. Rather than saying, in other words, that the Palace Hotel Ballroom gig was a consequence of Murphy's leaving his job tout court, we should say that the Palace Hotel Ballroom gig was instead a consequence of leaving his job *rather than* staying in his job (and not phoning up other guitar players).

If this is right, then there still remains a set of questions. Let's say that we hold that the Palace Hotel Ballroom gig is a consequence of Murphy's leaving his job relative to staying in his job and doing nothing else. OK, but why? To begin, it may be helpful to put this idea in the language of possible worlds. One possibility is simply to compare the world in which the given act ϕ was performed with worlds in which its alternative contrast is performed. Of course, this would be far too expansive, insofar as there are many worlds in which its contrast is performed, many of which will be very different and irrelevant for the purposes of this calculation. A world in which any time someone refuses to leave her job a nuclear explosion is triggered is surely not relevant for calculating the consequences of Matt "Guitar" Murphy's decision to leave his job rather than staying and doing nothing else. (He didn't, in other words, prevent a nuclear holocaust.) So we need to constrain the relevant *background conditions* against which the consequences of contrasting acts are ascertained.

Once the background conditions are identified, we may proceed as follows. If we rule out the act itself as counting among the act's consequences, it follows that the consequences of an act ϕ relative to an act ψ will be *differences in the states extrinsic to* ϕ and ψ, given the background conditions. But which extrinsic states? One possible answer—and, indeed, the one I accept—is *any*. Any difference in the states extrinsic to ϕ relative to ψ may be counted as a consequence of ϕ relative to ψ, and vice versa. This may seem expansive, but it delivers the right results. The Palace Hotel Ballroom gig is a consequence of Murphy's leaving his job relative to staying in his job and doing nothing, because the act in which he does nothing (given the relevant background conditions) does not bear an extrinsic relation to the Palace Hotel Ballroom gig, whereas the act in which he leaves his job to rejoin the Blues Brothers does. Because, given the assumption of the relevant act in

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

light of the relevant background conditions, that event does not take place, the gig does not exist. It is not a consequence relative to phoning another guitar player.

Thus, we may state this account of consequences in the following way:

> *Consequences*: a state of affairs S is a consequence of an act ϕ relative to an act ψ (performed at t) if S is a state of affairs extrinsic to ϕ, but not a state of affairs extrinsic to ψ, given a contextually identified set of background conditions.

(p. 106) This view has a number of salutary, well, consequences. To begin, *Consequences* makes it clear that the consequences of a given act relative to another one need not be identified causally. It is a consequence, for instance, of Barack Obama's failing to perform the safety check *relative to his performance of the safety check* that the plane crashes. Of course, Barack Obama's failure to perform the safety check did not cause it to crash. But it is nevertheless a consequence relative to his performance of the safety check. Second, because the relevant states are identified as *extrinsic*, the act itself will not be confused with the consequences of the act. The consequences of one act relative to another will always be identified in terms of objects, states, or events that are extrinsic to the act, or the relevant act description. Furthermore, *Consequences* can very easily be translated to account for the consequences of states of affairs, conditions, laws of nature, and so on. In each of those cases, we pick out a relevant contrast, and we identify the consequences of one set of laws of nature relative to another given the facts extrinsic to the relevant bearers of consequences.[19]

Of course, this proposal will face a number of objections. I consider them now.

5. Overinclusivity: Part I

Now, the suggestion that difference in states extrinsic to ϕ and ψ constitutes the consequences of ϕ relative to ψ may be thought too strong for (at least) two reasons. Problem one: some features of acts seem to be due to the act's extrinsic properties but in a way that should not be included in the act's consequences. For instance, let's say I rob a bank. The act of my robbing the bank surely has lots of consequences: I get rich and retire to Tahiti, the FDIC has to make a payout, and so on. But it doesn't seem right to say that one of the consequences of the act is the act's *illegality*. But it would also seem that the act's (p. 107) illegality is clearly a state extrinsic to the act. Given this, it would seem that my approach to an act's consequences is too expansive.

However, I think this criticism can be addressed. Recall that on this view the consequence relation will be assessed by comparing the extrinsic properties of two (or more) acts *given the relevant background conditions*. The latter phrase is key here. Consider my act of, rather than robbing the bank, sitting down and having a cup of coffee. Is the illegality of my act of robbing a bank at t a consequence of robbing the bank at t relative to my sitting down and having a cup of coffee? No. This is because the relevant extrinsic state—being performed in a world or context *in which robbing a bank at t is illegal*—is the

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

same for both acts. And so though illegality is an extrinsic property of my bank robbery, it is nevertheless not a consequence *given* that the context in which that extrinsic property is relevant is a world in which that illegality is held fixed.

6. Overinclusivity: Part II

Take “Guitar” Murphy. Imagine two possibilities. In the first case, he leaves his job with Aretha Franklin because he simply wants to play some great blues with the Blues Brothers: he’s a musician at heart, and he can’t stand being cooped up in a kitchen all day without playing guitar. So he jumps at the chance to rejoin the band and goes on to play the infamous gig at the Palace Hotel Ballroom. In the second case, he leaves not because he’s a diehard musician, but simply because he wishes to assist the Blues Brothers on their mission to save their childhood orphanage. In this case it would appear that the only extrinsic properties that differ between the first and second cases are the *motivations* of the act in question. But surely we would not wish to say that a *consequence* of the first case relative to the second is that “Guitar” Murphy is motivated to by love of music rather than beneficence toward an orphanage.

However, I think we need to be careful here. Begin with the act descriptions. How, in other words, are we distinguishing the aforementioned acts? Plausibly, the acts are (a) Murphy leaving as motivated by music and (b) Murphy leaving as motivated by beneficence. But if this is right, the consequences do not include the motives, insofar as the motives are part of the act descriptions. After all, the extrinsic properties of the actions involved *under these descriptions* are the same. But, of course, we may wish to describe the acts differently. We may wish to say, for instance, that the first case features the act of Murphy leaving the restaurant, which is, in turn, motivated by music. The second entails keeping the act description the same but varying only the motivation. And if this is right, it would appear that a consequence of the act under consideration relative to the other *includes* Murphy’s motivation—a difference in states extrinsic to the act.

However, we need to be even more careful here. What, exactly, is the *state* we are talking about when we are talking about Murphy’s motivation? One possibility is that we are talking about the state in which Murphy acted in accordance with motivation *x* (p. 108) rather than *y*. But if *this* is the state we’re interested in, it seems to me quite clear that it *is* a consequence of the act. And so there’s no problem.

But if “Murphy’s motivation” just refers to the state in which Murphy is motivated in way *x*, note that this state appears to me to permit of precisely the same treatment as the illegality of my potential bank robbery. Generally, when we’re interested in the consequences of Murphy leaving his job, we compare it to, more specifically, his *staying*, where his motivation (whether he loves music or is positively disposed toward orphans, etc.) is simply held fixed and, hence, is not a difference in the extrinsic properties of the act given the relevantly specified world within which the contrast takes place. To put this point another way, when assessing the consequences of a particular act, the motivations, state of mind, and so on of the agent, it is perfectly open given *Consequences* for the motivation in ques-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

tion to be treated as a part of the background conditions. And if this is right, there is no reason why *Consequences* need yield the problematic result, *especially* in conversational contexts in which we're interested in consequences *rather than* motivations.

However, there is a different problem in the same neighborhood that requires further address. To see it, imagine that if I go to law school at t, I will become a lawyer at t_{-1}. And also imagine that, if I go to med school at t, I will become a doctor at t_{-1}. Now, I will only become a lawyer if I go to law school. And I will only become a doctor if I go to med school. It seems right to say that one consequence of going to law school relative to going to med school is that I will become a lawyer. But, on the view I advocate, it would appear that one consequence of my *becoming a lawyer*, relative to becoming a doctor, is that I *go to law school*. After all, this is a difference in the states extrinsic to the contrast classes. But this is odd. It shouldn't be that my going to law school is a consequence of my becoming a lawyer, relative to becoming a doctor. The consequence relation between law school and lawyering should run in the other direction.

However, again, I think the problem here relies on a confusion in the contextually defined background conditions. Generally, in considering the consequences of my decision to become a lawyer, we consider alternative actions against the background of that which is already causally closed. I've already, in other words, gone to law school at t; I now have to determine the comparative consequences of becoming a lawyer at t_{-1} or becoming a doctor at t_{-1}, *given that I've gone to law school*. (Presumably, the consequences of becoming a lawyer will be much better.) However, if we wish to leave *open* the background conditions and, in other words, don't hold fixed that I've gone to law school at t, then it would appear perfectly sensible to say that *having gone to law school at t* is a consequence of becoming a lawyer relative to becoming a doctor at t_{-1} (which would entail having gone to med school at t). After all, it is a difference made to the world given that one becomes a lawyer rather than a doctor. Of course, given our interests in making relevant claims about consequences, we will generally wish to treat facts that are causally closed or simply determined at the time of action as part of the action's background conditions, and hence to hold that in the relevant comparison between doctoring and lawyering, we assume that one went to law school (which is what actually happened). (p. 109) But there's nothing—or so it seems to me—in the nature of the consequence relation *itself* that should rule out the alternative.[20]

7. Same Consequences

Let's imagine that I am standing at the counter in a coffee shop, and I can order precisely two possible drinks: a large (16 ounce) coffee and a small (12 ounce) coffee. No matter what I order, I will drink precisely 12 ounces of coffee. And so one would imagine that both actions have as a consequence that I drink 12 ounces of coffee. They have, to put this in a slightly different way, the same consequences relative to my coffee consumption. Of course, they will certainly have different consequences having to do with leftover cof-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

fee and so forth, but they will not differ when it comes to, for example, how caffeinated I am or how much actual coffee I consume.

But how can we say that these divergent actions have the same consequences in the way that seems natural here? After all, relative to, for example, my caffeine intake, my coffee consumption, and so on, both actions have the same extrinsic properties, and hence will not *differ* in their extrinsic properties along these lines. At most, one could say that the action of ordering the large coffee will have the consequence of some extra coffee being thrown away. But one cannot say of that action that it has the consequence that I drink 12 ounces of coffee, given that I drink 12 ounces of coffee in ordering the small coffee. (And vice versa: ordering the small coffee does not have the consequence of my drinking 12 ounces of coffee, given that it shares this extrinsic property with my action of ordering the large.) But this seems like a problem. Stated more abstractly: couldn't we say that two actions have the same consequences? But if consequences are crucially determined by reference to the difference in states extrinsic to competing acts, then this would seem impossible. In fact, it would seem as though any extrinsic states shared between two actions just aren't consequences of those actions at all.

In response, note that while it is true that my drinking 12 ounces of coffee is not a consequence of my ordering the small coffee *relative to ordering the large coffee*, it is nevertheless a consequence of my ordering the small coffee *relative to ordering no coffee at all*. Indeed, this is a consequence of both ordering the small coffee and ordering the large (p. 110) coffee relative to ordering no coffee at all. On the contrastive account I defend here, one must distinguish contrast classes (i.e., alternative actions) in determining what, precisely, the consequences of any given action ϕ are.

But understood in this way, the proposal I make is not at all strange. Indeed, in considering ordering the large or small coffee, we may be tempted to say things like: "when it comes to the amount of coffee I'll drink, it doesn't matter," "there's no difference between ordering the small and large, at least where coffee is concerned," and so on. But if there really is *no difference* between the two actions when it comes to the coffee I'll drink, it certainly seems right to say that my drinking 12 ounces of coffee is not a consequence of one relative to the other. Of course, these actions *do* have the same consequences when each action is considered in a pairwise comparison with, say, the action of ordering no coffee at all, or the action of ordering an espresso or a hot chocolate, and so on.

8. Conclusion: Is This Really What We Care about?

I began this paper by noting that consequences seem relevant in a number of normative domains and areas of philosophical concern. I then argued that the traditional view seems to get a number of results incorrect when it comes to the nature of consequences—the act (or, e.g., event and so forth) itself is not a part of the consequences of an act. For in-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

stance, it is quite true that the plane crash is a consequence of Barack Obama's failure to perform the safety check relative to his performance of the safety check. This is because the plane crash is a difference in the extrinsic property of his failure in comparison to his nonfailure. I then proposed an alternative, according to which the conceptual structure of the consequence relation is set given contextually specified contrasts, and the difference in extrinsic properties given these contrasts.

But is this what we really care about when we care about consequences? Indeed, this question could be put starkly: we appear to be interested in the metaphysical nature of consequences only insofar as we seem to care about them for normative purposes. But if we don't really care to assign a normative status to *all* the comparative extrinsic properties of a given act, why should we think that this account of consequences tracks what we're really interested in when we are, after all, interested in consequences? Do we really, in other words, care that a consequence of Barack Obama's failure to perform the safety check was a plane crash? Should that fact give us any bearing on the moral quality of Obama's actions?

Well, the answer here is either yes or no. And perhaps the answer is yes. Perhaps the plane's crash is relevant to our judgment about the moral quality of Barack Obama's actions, or the reasons he had to act or not act in this way. Now, to say this would not commit us to saying that Barack Obama behaved immorally in not preventing the plane crash. Perhaps there were stronger reasons that told against alternative courses of action. (p. 111) But if we care about the fact that Barack Obama's failure to perform the safety check led to the crash of the plane, then there is no barrier at all to saying that the current account of consequences matches up with what we, as a normative matter, seem to care about. Because this is something we care about. Indeed, it's important to remember that *Minimal Objective Consequentialism* concerns only *objective* reasons. If, in other words, had Barack Obama *known* that his failure to conduct the safety check would result in the plane's crash, would we have expected him to treat this as a moral reason significant for his decision making? Of course!

OK, maybe you're not convinced. Maybe we are disinclined to treat the fact that a consequence of Barack Obama's failure to perform the safety check was that the plane crashed, even an objective reason for Obama. This could be, among other potential possibilities, because Barack Obama had no reason to believe that a consequence of his failure to do so, relative to his doing so, was the crash of the airplane. Perhaps we may wish to hold Barack Obama only to account for the relevant rules that apply to the roles that he occupies (i.e., ex-President, father, and so on), and instead hold the airplane attendant (whose role it is to perform safety checks) morally responsible. And so forth.

But if we hold that we do not care about the plane crash in relation to judging the morality of Barack Obama's action, for these or other reasons, how should we describe our reticence? There seem to me at least two ways. First, we might attempt to preserve *Minimal Objective Consequentialism* and hold that:

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

1. *The plane's crash is not a consequence of Barack Obama's failure to perform the safety check, relative to his performance of the safety check.*

Alternatively, we might deny *Minimal Objective Consequentialism* and instead hold that:

2. *In evaluating the morality of Barack Obama's actions at the relevant time, we do not judge as relevant all of the consequences, but instead (at most) a subset of such consequences.*

Consequences is compatible with (2), but not (1). But, or so it seems to me, (2) is no less defensible in and of itself than (1), and hence if there is good reason to accept *Consequences*, independently of our commitment to evaluating the morality of Barack Obama's actions in terms of its consequences, then there should be no barrier to doing so insofar as (2) is just as defensible as (1). If so, whether we should accept *Consequences* will not await an inquiry into the normative significance of consequences, but rather vice versa.

References

Bennett, J. 1995. *The Act Itself*. Oxford: Oxford University Press.

Bergström, L. 1966. *The Alternatives and Consequences of Actions*. Stockholm: Almqvist and Wiksell.

(p. 112) Bernstein, S. 2014. "Omissions as Possibilities." *Philosophical Studies* 167: 1–23.

Brandt, R. 1959. *Ethical Theory*. Englewood Cliffs, NJ: Prentice Hall.

Dorsey, D. Unpublished manuscript. "On Prudence."

Feldman, F. 1986. *Doing the Best We Can*. Dordrecht, the Netherlands: D. Reidel.

Haines, W. 2019. *Consequentialism*. *Internet Encyclopedia of Philosophy*, ISSN 2161–0002, **https://www.iep.utm.edu/**.

Hart, H. L. A., and Honoré, A. M. 1959. *Causation in the Law*. Oxford: Oxford University Press.

Henne, P., Pinillos, Á., and De Brigard, F. 2016. "Causation by Omission and Norm." *Australasian Journal of Philosophy* 95: 270–283.

Kamm, F. 2007. *Intricate Ethics*. Oxford: Oxford University Press.

Lenman, J. 2000. "Consequentialism and Cluelessness." *Philosophy and Public Affairs* 29.

McGrath, S. 2005. "Causation by Omission: A Dilemma." *Philosophical Studies* 123.

Moore, G. E. 1952. "A Reply to My Critics." In *The Philosophy of G. E. Moore*, edited by P. Schilpp. New York: Tudor, pp. 627–667.

Moore, G. E. 2005. *Ethics*. Edited by W. Shaw. Oxford: Oxford University Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Raphael D. D. 1956. “The Consequences of Actions.” In *Aristotelian Society Proceedings*. London: Harrison and Sons, pp. 91–119.

van Inwagen, P. 1983. *An Essay on Free Will*. Oxford: Oxford University Press.

Notes:

([1]) This character is not identified by name. (Indeed, the script just calls her “Waitress.”) However, given that most of the characters in the film play themselves, I don’t think it’s out of bounds to identify her as Aretha Franklin.

([2]) See, for instance, Dorsey, unpublished manuscript.

([3]) Raphael (1956), 106. Now, lest his view be confused with a purely causal account (see later), Raphael suggests that “To determine the consequences of an action is to make a practical not a factual judgment” (Raphael 1956, 106).

([4]) Hart and Honoré (1959), 65–66.

([5]) Cf. Lenman (2000), 344–345.

([6]) Moore (1952), 559.

([7]) Notice that Fred Feldman, in his development of a utilitarian view, holds that the essential relation is not the consequence relation, but the “accessibility” relation, where moral agents are required to bring about the best world that is accessible to them. Accessibility, here, clearly identifies the world that would be brought about given the action. Cf. Feldman (1986), 36.

([8]) Cf. Bergström (1966), 66. See also Moore (2005), 7.

([9]) “In consequentialism, the ‘consequences’ of an action include (a) the action itself, and (b) everything the action causes” (Haines 2019). See also Brandt (1959), 354n2.

([10]) Though I will not discuss his view here, Jonathan Bennett develops the notion of a noncausal consequence in Bennett (1995), 39. For Bennett, noncausal consequences are facts of entailment between true propositions (Bennett 1995, 41). However, it seems to me that noncausal consequences go beyond simply these cases.

([11]) Kamm (2007), 140–141. My emphasis.

([12]) Cf. McGrath (2005), 125.

([13]) Bernstein (2014), 2–3.

([14]) Cf. Bernstein (2014); Henne, Pinillos, and De Brigard (2017); McGrath (2005); and others.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([15]) Note that this holds even if the latter constitutes the former, as constitution is not a reflexive relation.

([16]) Bennett (1995), 43.

([17]) Bennett goes on to consider accounts of “intrinsic” properties of acts that include all the *simultaneous* facts about an action, including intrinsic and extrinsic properties. He finds this wanting, however. See Bennett (1995), 44.

([18]) van Inwagen (1983), 16.

([19]) We may wonder how to evaluate consequences relative to each other on this view. Indeed, perhaps given our moral or normative proclivities, we may be quite interested in what it would be for one act ϕ to have the *best* consequences. I don’t mean by this the general axiological question, that is, how we rank-order sets of consequences (i.e., whether we should, for instance, calculate the comparative degree of pleasure, or desire satisfaction, or some other thing). Rather, I ask a somewhat more technical question: how, given such an axiology, do we generate an acceptable rank ordering of acts given their consequences (relative, presumably, to each other)? I think the answer should be relatively simple. Consequences are rank-ordered given pairwise comparisons with other acts. Let’s imagine that there are three acts, ϕ, ψ, and π. If I ϕ, three units of pleasure will be generated, if I ψ, two, and so on. How are these acts to be rank-ordered? Well, consider the pairwise comparison of ψ and π. Here it would appear (assuming that pleasure is valuable) that ψ is better than π—the consequences of ψ relative to π are better than the consequences of π relative to ψ. The consequences of ϕ relative toψ are also better than the consequences of ψ relative to ϕ. And hence we should rank ϕ as better than ψ. By transitivity, then, we have a rank ordering of $\phi > \psi > \pi$. And this is borne out in further pairwise comparisons. The consequences of ϕ *relative to both alternatives* are better than the consequences of both alternatives relative to ϕ. And hence ϕ is the action among my alternatives which has the best consequences.

([20]) One final problem arises given this example. Recall that, or so I argued, the consequence relation should be nonreflexive. But going to law school and becoming a lawyer are both states that are extrinsic to each other. And if this is right, it would seem sensible to say that they are both consequences of the other. However, this is not accurate given the account stated. If we allow that going to law school is a consequence of becoming a lawyer, it is only in contrast to becoming a doctor at t_1. If we allow that becoming a lawyer is a consequence of going to law school, it is only in contrast to going to med school at t. And so the consequence relations here are different: they permit of different contrast classes. (To put this another way, as *Consequences* makes clear, the consequence relation fixes the contrast to a state or act that happened at the time of the act for which we seek to identify the consequences.)

Dale Dorsey

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Dale Dorsey is Dean’s Professor and Chair of the Department of Philosophy at the University of Kansas. He generally works in normative ethics, at the intersection of the personal good, morality, and practical rationality. He has also worked on metaethics and has written essays on the moral philosophy of David Hume, Francis Hutcheson, and John Stuart Mill.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Alternatives

Holly M. Smith

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.2

Abstract and Keywords

Consequentialists have long debated (as deontologists should) how to define an agent's alternatives, given that (a) at any particular time an agent performs numerous "versions" of actions, (b) an agent may perform several independent co-temporal actions, and (c) an agent may perform sequences of actions. We need a robust theory of human action to provide an account of alternatives that avoids previously debated problems. After outlining Alvin Goldman's action theory (which takes a fine-grained approach to act individuation) and showing that the agent's alternatives must remain invariant across different normative theories, I address issue (a) by arguing that an alternative for an agent at a time is an entire "act tree" performable by her, rather than any individual act token. I argue further that both tokens and trees must possess moral properties, and I suggest principles governing how these are inherited among trees and tokens. These proposals open a path for future work addressing issues (b) and (c).

Keywords: action theory, act sequence, act token, act tree, act versions, alternative, Alvin Goldman, consequentialism, co-temporal acts, inheritance principle

1. Introduction

MOST moral theories require or forbid an action in light of how it compares to its alternatives. Thus a canonical statement of consequentialism's principle of obligation might be expressed as follows (Carlson 1999, 253):

> CO: An action ought to be performed if and only if its outcome is better than that of every alternative.

Similarly, many deontological moral theories hold that whether an action ought to be done depends on whether the agent's performing it would better fulfill her duties than performing any alternative. This view could be expressed as follows:

> DO: An action ought all-things-considered (ATC) to be performed if and only if its net fulfillment of pro tanto duties is superior to that of every alternative.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

In both consequentialism and deontology the concept of an action's alternatives is key: the action's ATC moral status depends on how its morally relevant features compare to the morally relevant features of the agent's alternatives. Unfortunately, as early work by consequentialists revealed, it turns out to be far from trivial to identify what counts as an alternative for an agent, or what entities her alternative set includes (Bergström 1966; 1971; Castañeda 1968). Three such problems are especially prominent.

1.1. Multiple Versions of Acts

According to many theorists, any particular action has many different "versions."[1] Thus trial witness Rick, asked whether the defendant shot the victim, says "no." In saying "no" (p. 114) Rick tells a lie, relieves the anxiety of the defendant's spouse, and secures the acquittal of the defendant. These are different "versions" of his act. The causal consequences of these versions may be different, as may their other morally significant properties. Rick's saying "no" causes the defendant's acquittal, but his relieving the spouse's anxiety does not cause the defendant's acquittal. Rick's telling a lie violates a deontological duty, but his relieving the spouse's anxiety does not violate such a duty. This gives rise to several problems. First, depending on which of these act versions is selected as one of Rick's alternatives, we may obtain a different moral evaluation of what he did.[2] We need a nonarbitrary way to make this selection.

Second, depending on which alternative sets are allowed, and which principles of deontic logic are accepted, it can be argued that the existence of multiple versions of an act leads to contradictory prescriptions.[3] For a short version of this claim, consider Sam, who is choosing between giving the king a poisoned drink (which would be wrong) or giving him a healthy drink (which would be obligatory). His giving the king a healthy drink entails another version of this act, namely his giving the king a drink. According to a common principle of deontic logic, if act X entails act Y, then X's being obligatory entails Y's being obligatory. Since Sam's giving the king a healthy drink entails his giving the king a drink, his giving the king a drink is also obligatory. But suppose it's true that if Sam gave the king a drink, he would give the king a *poisoned* drink. Then the consequences of Sam's giving the king a drink would be disastrous, so his giving the king a drink is wrong. But then Sam's giving the king a drink is both obligatory and wrong![4]

1.2. Co-temporal Acts

Sometimes an agent can perform simultaneous acts that are independent of each other rather than versions of each other. For example, double agent Veneeta has the option of scratching her head with her right hand (thus signaling to her confederates), while also scratching her hip with her left hand (thus signaling to the CIA). Call these "co-temporal" acts.[5] We can view Veneeta as having two independent alternative sets: (1) scratching her head or not scratching her head, and (2) scratching her hip or not scratching her hip. Or we can view her as having a single set of four alternative co-temporal acts: (A) scratching her head while scratching her hip, (B) scratching her head while not scratching her hip, (C) not scratching her head while scratching her hip, and (D) not scratching her head

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

while not scratching her hip. What the consequences are of her scratching her head may depend on whether she simultaneously scratches her hip (and vice versa). Given certain assumptions about the consequences and interactions of these various (p. 115) acts, which alternative set we choose will generate different prescriptions. Moreover, given certain plausible deontic principles, it is possible to derive serious problems for either of these rival sets of alternatives.[6]

1.3. Sequences of Acts

Similar issues arise when we ask how to identify an agent's alternative set when the agent has the option of performing various possible sequences or courses of action. For example, Veneeta may have the option of scratching her head and then rubbing her nose (thus sending one signal to her confederates). Or she might scratch her head and then not rub her nose (sending a different signal), or she might not scratch her head but then rub her nose (sending still a third message), or not scratch her head and then not rub her nose (sending a fourth message). Suppose it's true that if Veneeta scratched her head she would then rub her nose. Should we say that her alternative set nonetheless includes the sequence scratching her head and not rubbing her nose (a sequence she would not in fact perform)? How we resolve the question of what sequences should be included among her alternative sequences can make a major difference to what she is morally required to do and whether these requirements are intuitively acceptable.[7]

The three problems just outlined arise because of the structure of human action.[8] To solve these problems we need to examine this structure, a task that prior investigators have not addressed with satisfactory depth. The aim of this chapter is to sketch a richer and more robust account of human action, utilizing a fine-grained approach, that enables us to resolve the problem of multiple act versions. Addressing how to identify the proper alternative sets in the case of co-temporal actions and act sequences exceeds the space limitations on this chapter. However, these problems also crop up in the debate between Actualism and Possibilism (see Cohen and Timmerman, Chapter 7, this volume). This chapter will rest content with providing foundational tools for resolving these questions.

2. Constraints on Solutions

How to define an agent's alternatives has been discussed primarily by consequentialists, led by the seminal work of Lars Bergström (1966) and Hector-Neri Castaneda (1968). However, (p. 116) it is clear that the question arises with equal or even greater force for deontological theories.[9] As discussions of Kant's theory have noted, for example, an agent's act may be described as "getting some money from the bank" or described with equal accuracy as "robbing the bank" (Glasgow 2012). Presumably a maxim using the latter description would fail the Categorical Imperative test and so be wrong, while arguably one using the former description would pass the Categorical Imperative test, and so may be permissible.[10]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

In this chapter I will briefly outline a robust theory of human action, describe the contemporary debate about act versions in the literature (expanded to encompass deontology in addition to consequentialism), and then sketch my own solutions. I begin by proposing several plausible constraints on any acceptable solution.

> **Constraint A:** An action counts as an alternative for an agent only if that action is a concrete act (as contrasted with an act type) and is performable by the agent. A set of actions counts as an alternative set for an agent only if the set includes concrete acts that are performable by the agent, time-identical (that is, either the time-intervals of each act are the same, or at least start at the same time);[11] mutually exclusive, and jointly exhaustive.
>
> **Constraint B:** An account of an agent's alternative set must be normative-theory neutral: the same set of actions is identified as the agent's alternative set in the context of every plausible normative theory.

Constraint A captures common ground assumed by many theorists.[12] In Constraint B, "normative theories" include rival first-level moral theories, such as consequentialism and deontology.[13] But they also include nonmoral accounts of what is normatively required or forbidden. A constraint like Constraint B has not, to my knowledge, been discussed before. However, it is clearly important for comparative evaluations of rival moral theories, or of morality as against rationality or prudence. For suppose, in the context of consequentialism, an agent is deemed to have an alternative set consisting of actions ***x***, ***y***, and ***z***, each performable at time t_1. However, in the context of a deontological theory, the agent is deemed to have an alternative set consisting of actions ***m***, ***y***, and ***n***, each also performable at time t_1. Furthermore, consequentialism implies that the agent ought to perform ***x***, while deontology implies that the agent ought to perform ***m***. (p. 117) Given these recommendations, we lack one of our chief ways for comparing the concrete prescriptions of consequentialism and deontology in order to judge which is most acceptable. The same difficulty can arise if morality identifies one array of alternatives, but prudence or rationality identifies a different array of alternatives on the same occasion. If these alternative sets have different members, we have lost one of our most important tools for determining whether it is best to follow the recommendation of morality or that of prudence.[14]

3. The Structure of Human Action

Let us start our exploration of the structure of human action by examining how to deal with the fact that at a given time an agent performs multiple (but closely related) actions, some with different consequences or other morally relevant features than others. We'll focus here on what an agent does when she acts, not on defining her alternatives. Consider *Donation*:

> *Donation*:

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> Last week Lisa promised her son Charles, who is gay, never to support any organizations that work against gay rights. At 1:00 today, as part of her year-end charitable donations, Lisa donates $1,000 to Advance Charities by clicking the “Donate” button on its website. Advance Charities, an organization that raises funds and distributes them to various specialized charities, funnels some of its donations to UNICEF. Through this process Lisa’s contribution saves five lives. However, unknown to Lisa, Advance also supports Preserve Marriage, an organization that funds legal defenses for businesses that refuse to serve gay married couples. Lisa’s donation enables Preserve Marriage to successfully defend a wedding planner who refuses to plan an upcoming gay wedding. The wedding is thrown into disarray and must be rescheduled. The engaged couple sustains substantial legal costs and loses several thousand dollars in deposits with other wedding vendors. Through Advance, Lisa’s donation supports Preserve Marriage, and so violates her promise to Charles.

At 1:00 Lisa performs a number of what are often called act “versions.” For example, she moves her finger; she presses the “Donate” button; she donates to Advance Charities; she donates to UNICEF; she saves five lives; she donates to Preserve Marriage; she causes the gay couple to lose a substantial amount of money; she supports an organization that works against gay rights; and she breaks her promise to Charles. She performs some of these acts *by* performing others: for example, she presses the “Donate” button by moving (p. 118) her finger; she donates to Advance by pressing the “Donate” button; she donates to Preserve Marriage by donating to Advance; she causes the gay couple to lose a substantial amount of money by donating to Preserve Marriage; she breaks her promise by donating to Preserve Marriage; and she saves five lives by donating to UNICEF. However, not all these actions are directly related to each other through the “by” relation. For example, Lisa doesn’t donate to UNICEF by donating to Preserve Marriage (or vice versa). And she doesn’t break her promise by donating to UNICEF. In these cases, both acts are done by performing a third act, for example, moving her finger.

Some action theorists characterize Lisa’s situation as one in which only a single action is performed, but in which the action can be described in a variety of different ways, such as “moving her finger” and “donating to Advance” (classically, Anscombe 1958; Davidson 1963). Other action theorists have advocated a more fine-grained view of act individuation and characterize Lisa’s situation as one in which she performs a great many distinct actions, although ones that have a special close relationship to each other (most prominently, A. Goldman 1970). For our purposes it will be illuminating to adopt the more fine-grained view.[15]

On Goldman’s theory, an act token is an event consisting of the exemplification of an act property (such as *moving one’s finger* or *donating to Advance*) by an agent at a time.[16] We will restrict our attention to possible act tokens that are performable by the agent in question.[17] When an agent performs a basic act (usually an intentional bodily movement), that act *generates* certain higher-level acts.[18] We refer to this when we say that the agent performs the higher-level act *by* performing the lower-level act. Goldman provisionally

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

distinguishes four different kinds of level generation: causal generation, conventional generation, simple generation, and augmentation generation. These are defined as follows:

- In *causal generation*, the lower-level act ***a*** causes a certain effect E, and because of this effect, the agent performs the upper-level causally generated act ***a****. Thus Lisa's donating to Preserve Marriage causes the couple to lose thousands of dollars, so her action of donating to Preserve Marriage causally generates another action, her causing the couple to lose money.

(p. 119) • In *conventional generation*, a lower-level act ***a*** occurs in certain circumstances C, and the occurrence of the action in these circumstances, together with existence of a certain rule R saying that ***a*** done in C counts as act ***a****, guarantees the performance of ***a****. Lisa's breaking her promise to Charles, in circumstances where there is a moral rule that breaking promises is pro tanto wrong, generates by conventional generation her act of doing something pro tanto wrong.

- In *simple generation*, a lower act ***a*** occurs in circumstances C, and the act together with the circumstances ensures the performance of act ***a****. Thus Lisa's donating to Preserve Marriage, done in circumstances in which she has promised not to support anti-gay organizations, generates by simple generation her act of violating her promise.[19]

- *Augmentation generation* is similar to simple generation, but with the distinction that performance of the upper-level act ***a**** "entails" performance of the generating act ***a***.[20] The generated act arises from the "augmentation" of the generating act by being performed in some manner or in some circumstance. Thus Lisa's act of moving her finger generates, by augmentation generation, her act of moving her finger slowly. In such cases the occurrence of the generated act guarantees that the generating act occurred as well. Goldman notes that augmentation generation seems rather different from the other three forms of generation, not least because an act ***a**** that is generated by an act ***a*** via augmentation generation is not felicitously described as being performed "by" the performance of ***a.*** We wouldn't normally say that Lisa moved her finger slowly by moving her finger. Nonetheless Goldman defends the claim that pairs of acts are linked as described in this manner by arguing that this view seems superior to alternative descriptions of their relationship. In particular, it makes more sense to say that Lisa's moving her finger generated her moving her finger slowly than to say that her moving her finger slowly generated her moving her finger (and we certainly wouldn't say her moving her finger was done by her moving her finger slowly).[21] In every case of generation, the generated act occurs because of some additional fact or event not encompassed by the generating act. This feature is preserved in augmentation generation when we hold that Lisa's moving her finger generated her moving her finger slowly, but it would be violated by our holding that Lisa's moving her finger slowly generates her moving her finger.

(p. 120) Several further aspects of Goldman's fine-grained theory should be mentioned. First, generation is transitive, so that a lower-level act ***a*** may generate, through a series of (possibly different) types of generation, an upper-level act ***a****. Thus Lisa's moving her

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

finger generates her donating to Preserve Marriage, which in turn generates her causing the engaged couple to lose a good deal of money. So her moving her finger also generates her causing the engaged couple to lose money. Second, two acts can each be generated by a third act when neither of the two generates the other. Lisa's act of donating to Advance generates her saving five lives, and also generates her causing the engaged couple to lose money, but her saving five lives doesn't generate her causing the couple to lose money (or vice versa). They are generated on diverging branches stemming from a common generating act. All Lisa's acts clearly have a close relationship with each other, even though none is identical with any of the others. Goldman would diagram their relationships as in Figure 6.1.

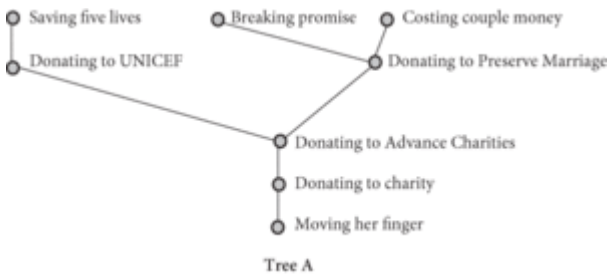


Figure 6.1. Lisa's act tree

This is (part of) what Goldman terms an "act tree."[22] It illustrates the structural relationships among acts that make up what Lisa does when she moves her finger and presses the "Donate" button. Of course, there are many more acts on this act tree than those displayed here; this only provides a partial picture of how all these acts are related to each other to form an important cluster. Any act tree has an infinite number of acts on it. Such an act tree roughly corresponds to what coarse-grained theorists call "a single act."

According to the fine-grained view, one salient feature of different act tokens, even though they are members of the same act tree, is that they can have different properties from each other.[23] For example, Lisa's act of donating to Preserve Marriage has the property of *causing the engaged couple to lose money*, but her act of donating to UNICEF does not have the property of *causing the engaged couple to lose money*. Of special interest is the fact that different acts on the same act tree can (directly) have different moral (p. 121) properties. Thus Lisa's act of saving five lives may have the moral property of *being pro tanto right*, while her act of breaking her promise has the contrasting moral property of *being pro tanto wrong*. Note that an act having a moral property, such as the property of *being pro tanto wrong*, generates (via conventional generation) a separate act of doing something pro tanto wrong. The act of doing something pro tanto wrong does not itself have the property of *being pro tanto wrong*. For another example of two acts on the same tree having different moral properties, suppose Lisa and her husband have agreed that their year-end donations will not exceed $500. However, to spite her husband, Lisa gives more to Advance than the agreed-on sum. Then her donating $1,000 to Advance has the property of *being blameworthy*.[24] Nonetheless, her giving to Preserve Marriage is not blameworthy, even though it is at least pro tanto wrong: she is not aware (and, let's stipu-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

late, could not have foreseen) that she performs this act. Suppose duty requires donating a moderate percentage of one's income to charity, $500 in Lisa's case. Then her act of donating $1,000 has the property of *being supererogatory*, while her act of donating to charity is not supererogatory.

When I say that the act of keeping a promise is pro tanto obligatory, but the act of giving more than a moderate amount to charity is not pro tanto obligatory, I am relying on the assumption that an act token has a given moral property directly in virtue of the act type of which it is a token.[25] Moral properties directly adhere to act tokens in virtue of their act types. But it is sometimes claimed that an act token can possess a moral property indirectly—can *inherit* that property—in virtue of its relation to another act token on the same act tree.[26] Thus it might be claimed that Lisa's donating to Advance inherits pro tanto wrongness by virtue of the fact that it generates a higher-level act, namely her breaking a promise, which has this property directly in virtue of its act type. If this inheritance thesis is correct, it will be incumbent on its advocates to explain in virtue of which relationships an act can inherit a given moral property. If act token ***a**** has moral property *P*, does every act that generates ***a**** inherit *P*? Does every act that ***a**** generates also inherit *P*? Does every act on the same act tree as ***a****, whether or not it generates or is generated by ***a****, inherit *P*? And does this vary depending on the moral property in question, so that (say) if ***a**** has the property of *being pro tanto obligatory*, then all acts that ***a**** generates inherit the property of *being pro tanto obligatory*, but if ***a**** has the property of *being blameworthy*, then some acts that it generates fail to inherit this property (since some of these acts may be unintentional)? Prior scholars writing on alternatives may have overlooked the fact that on the same act tree there can be pairs of acts in which neither act immediately or distantly generates the other. If one of these acts is morally (p. 122) significant, the assumption that inheritance depends on generation tells us nothing about the moral significance of the second act, which neither generates nor is generated by the first act. Settling inheritance questions is not an easy task, and it is complicated by the fact that many nonmoral properties possessed by acts on the same act tree are not inherited. For example, the fact that Lisa's donating to Advance is intentional does not mean that her giving to Preserve Marriage inherits that intentionality, and the fact that her giving to Preserve Marriage causes the couple to lose money does not mean that her donating to UNICEF causes the couple to lose money.

4. Three Proposals for Identifying an Agent's Alternatives

Let us return to our main query. In light of the fine-grained analysis of act individuation, what kind of entity best serves as one of the agent's alternatives (or, equivalently, as a member of the agent's set of alternatives)?

We require an entity having several features. First, it should have all the properties of the agent's conduct that are morally relevant to the moral theory we are considering. Setting aside the possibility of inherited moral properties, then Lisa's act of donating to UNICEF

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(see Figure 6.1) would not be the right act to identify as one of her alternatives when we are considering consequentialism. (The acts to which this is an alternative would be components of other act trees, which are not depicted in Figure 6.1.) Although this act has some morally important causal consequences (saving five lives), it doesn't have *all* the morally important causal consequences of Lisa's conduct as a whole (for example, it doesn't have the consequence of causing the engaged couple to lose money). But second, as is implied by Constraint B, the act should have all the properties that are relevant to *all plausible normative theories*, not just the properties relevant to our target moral theory. Only insofar as this is true can we compare the implications of our target moral theory to those of rival theories and determine which theory issues the most acceptable prescriptions. Lisa's act of donating to UNICEF also fails this second test, because it doesn't have the properties of breaking a promise or violating a pro tanto duty (possessed by some acts on the same tree), properties which aren't relevant to consequentialism but are relevant to many deontological theories.

Given these considerations, three proposals regarding which act should be identified as a member of the agent's alternative set naturally suggest themselves. Proposal I selects the bottom-most basic act on a tree as one of the agent's set of alternatives. Proposal II selects, as one of the agent's set of alternatives, the highest act on a tree which is such that all morally significant acts on the tree are either identical to, or generated by, this act.[27] (p. 123) Proposal III selects, as one of the agent's alternatives, an entire performable act tree. Let's examine these in turn.

4.1. Proposal I: Bottom-most Acts as Alternatives

The first proposal is that the "bottom-most" act (Bottom-most, for short)—a simple bodily movement—serves as one of the agent's alternatives. **Bottom-most** has plausibility because this act generates all the upper-level acts on the same act tree, and—especially significant for consequentialism—it also causes all the consequences traceable to the agent's conduct on this occasion. If Lisa's moving her finger is the bottom-most act on her act tree, then it generates all the higher-level acts, and it also causes five lives to continue, the couple to lose money, and relations between Lisa and Charles to be strained. The values and disvalues of the causal consequences can all be ascribed to Lisa's moving her finger. For a consequentialist, the values of these consequences can then be compared to the values of the consequences of other bottom-most acts Lisa could instead perform. These comparisons would enable us to identify the all-things-considered (ATC) moral status of her moving her finger. Moreover, if (and only if) we assume that the moral properties of higher-level acts are inherited by lower-level acts that generate them, then the bottom-most act inherits the deontologically relevant properties possessed by all the acts it generates. Thus a deontologist could claim that the bottom-most act inherits the deontologically significant properties of *being pro tanto obligatory*, *being pro tanto wrong*, *violating a right*, and so forth. In this manner the deontic values and disvalues of Lisa's deontologically significant upper-level acts could also be ascribed to her bottom-most act and compared with the parallel deontic values that would accrue to alternative bottom-most acts if she did something else instead. Her bottom-most act, such as moving her fin-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ger, would be identified as her alternative by both consequentialism and deontology, as well as by other normative theories, thus apparently satisfying Constraint B.[28]

However, there are several key defects with **Bottom-most**.

(1) First, when we select the agent's "bottom-most" act, relative to which set of acts are we determining that it is bottom-most? This set of acts can't be all the possible acts that Lisa could perform at that time, since the target act doesn't cause the consequences of, or generate, some of the other acts she could perform instead, such as her donating $500 rather than $1,000 to UNICEF. The relevant set of acts is implied in our description of **Bottom-most**: it's the set of acts *on the same act tree* as the target act. Which act counts as bottom-most, and what its properties are, depends on the act tree of which it is a member. But this suggests that it is (p. 124) fundamentally the act tree, rather than the bottom-most act, that functions as the agent's alternative.

(2) A second key defect with **Bottom-most** is that an agent may have what we would intuitively recognize as two or more alternatives in which the bottom-most acts are apparently the same. For example, suppose Lisa is looking at Advance's donations webpage, on which there are boxes to be checked for various levels of donations. She can tap the box for giving $100, or the box for giving $500, or the box for giving $1,000. Whichever box she taps, her activity will have a bottom-most act of *moving her finger.* Of course, to tap the box for giving $500, she will have to move her finger in a different way (let's call this *moving her finger in way* Y) than she would have to move it in order to tap the box for giving $1,000 (let's call this moving her finger in way X). Lisa's moving her finger generates, by augmentation generation, her moving her finger in way X (or her moving her finger in way Y). So the token of moving her finger, bottom-most in one act tree, might also appear as the bottom-most token in some of her other performable act trees, in cases where we want to separate these as distinct alternatives for the agent. In Lisa's case, moving her finger could generate her donating $100, or her donating $500, or her donating $1,000. **Bottom-most** appears to violate the spirit of Constraint A, because it fails to recognize different alternatives as distinct and mutually exclusive. Here again, it is the act trees that seem to be occupying the position of alternatives.

A supporter of **Bottom-most** could reply that although Lisa's moving her finger so as to donate $500 and her moving her finger so as to donate $1,000 are indeed acts that are agent-identical, time-identical, and act type-identical, they are nonetheless not identical act tokens, since they are members of different possible act trees.[29] Indeed, referring to these tokens by the same name is likely to mislead us, as we will see in Section 5.2. We could relabel the first token "moving-her-finger$_Y$" and the second one "moving-her-finger$_X$," thus defusing the appearance that these tokens cannot be alternatives to each other. However, incorporating this into **Bottom-most** makes it highly cumbersome to refer unambiguously to the alternatives we mean to identify, and such a reference must refer to the act tree of which the token is a component in order to pick out the act token in question. So in this second way **Bottom-most** again depends critically on the act tree of

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

which the bottom-most act is a member. In light of this it appears wise to seek a better proposal.

4.2. Proposal II: Highest Normatively Significant Act Tokens as Alternatives (Highest)

Highest says that an agent's alternative should be understood as the highest act which is such that all normatively significant acts on the tree are either identical to, or generated (p. 125) by, this act. In assessing **Highest** it will be helpful to shift our focus from the normative property an act possesses to the normatively significant act type of which it is a token. Then **Highest** says that the agent's alternative should be understood as the highest act such that all acts on the same tree that are tokens of normatively significant act types are either identical to, or are generated by, this act. This proposal aims to evade the second difficulty with **Bottom-most**, since Lisa's bottom-most act, moving her finger, is *not* the highest one that is identical to, or generates, all the normatively significant acts she performs. Donating to Advance is higher than her bottom-most act of moving her finger, but this higher act also generates all her normatively significant acts, so under **Highest** moving her finger could not qualify as her alternative.

To fill out **Highest**, we need a more precise specification of a "normatively significant act." Let us adopt the view that normatively significant acts include any act token of a right-making act type, such as *breaking a promise* or *producing consequences with greater utility than any alternative act*. They should also include any act of a normative act type, such as *violating a duty* or *doing something ATC wrong*.

Let's assume that in Lisa's case, as viewed by a deontologist, her morally significant acts include breaking a promise (to her son), breaking a promise (to her husband), saving five lives, two acts of doing something pro tanto wrong, one act of doing something pro tanto right, and (let's assume) fulfilling a smaller net balance of pro tanto duties than some alternative, and doing something ATC wrong. As viewed by a consequentialist, her morally significant acts include (let's assume) producing consequences with less utility than some alternative act and doing something ATC wrong. To implement **Highest**, we must identify which act on her act tree is the highest one identical to, or generating, all of these morally significant acts. Inspection of Figure 6.1 suggests that Lisa's donating to Advance is the highest-level act that generates all of the acts that are deontologically or consequentially significant. On **Highest** this act qualifies as one of Lisa's set of alternatives, which is comprised of all the highest morally significant acts on the act trees performable by Lisa on that occasion.

But **Highest**, like **Bottom-most**, suffers important difficulties.

(1) Alas, the first difficulty is identical to the first difficulty with **Bottom-most**. To accurately define the relevant set of acts among which a target act is the highest one that is identical to, or generates, all the normatively significant acts an agent performs, we must clarify that all the latter acts are the members of the act tree of which the target act is it-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

self a member. This again suggests that it is the act tree, rather than the highest morally significant act token on that tree, that functions as the agent's alternative.

(2) The second problem is identical to the second problem with **Bottom-most**. Suppose a deontological theory recognizes *giving to charity* as an obligation-making act type: any token of this act type immediately generates the agent's doing something pro tanto obligatory. Within such a theory we cannot identify Lisa's alternative as her donating to Advance Charities (as we did earlier), because donating to Advance Charities is higher than giving to charity (a morally significant act), and so doesn't qualify as the (p. 126) highest act on the act tree such that all morally significant acts on the tree are either identical to, or generated by, this act.[30] We must instead identify Lisa's giving to charity as one of her (highest normatively significant) alternatives. But this has the same problem we encountered when we applied **Bottom-most** to Lisa's case. Lisa has what we intuitively recognize as several alternatives, including donating $500 to Advance. But this act, like donating $1,000 to Advance, would also be generated by an act token of giving to charity, which is morally significant according to our present deontological theory. Intuitively we believe that Lisa has at least two alternatives—donating $500 to Advance and donating $1,000 to Advance. But according to **Highest**, these apparently comprise only *one* alternative, giving to charity. This doesn't allow us to perspicuously differentiate these alternatives as two rather than one. It's important in the context of our deontological theory to keep them distinct, because Lisa's donating $1,000 to Advance is morally quite different from her donating only $500 (the larger donation saves more lives, but also violates her promise to her husband).

We are forced to conclude, then, that **Highest**, like **Bottom-most**, appears to violate the spirit of Constraint A, because it too fails to recognize distinct alternatives as distinct and mutually exclusive. Here again, we might attempt to elude this problem by recognizing two distinct act tokens, giving-to-charity$_A$ and giving-to-charity$_B$. But as we saw before, this would make it highly cumbersome to refer unambiguously to the alternatives we mean to identify, it could not be done without referring to the trees in question, and in any case fails to resolve the first problem.

(3) Finally, those who believe that normative theories must offer usable decision guides to agents will notice that **Highest** suffers from a severe practical difficulty.[31] Unless one has a list of all the plausible normative theories (which are probably infinite in number), and therefore a list of all the plausible normatively significant act types, it isn't possible to identify which act is identical to, or generates, all the acts on a given tree that are tokens of normatively significant act types. Since no human being possesses such a list, no one is in a position to correctly identify an agent's set of genuine alternatives. To correctly identify some act token ***a*** as an alternative, one would have to ascertain that no lower-level token that generates ***a*** is of a normatively significant act type according to any of the infinite number of plausible normative theories, surely an impossible task.[32] Of course, there are other limitations to knowledge about an agent's alternatives. Anyone who lacks full knowledge about what acts an agent is physically able to perform will not be able to determine exactly what alternatives are available to the agent. Similarly, even if we pos-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

sessed a list of all normatively significant act types, ascertaining the *actual normative* status of an agent's possible act tokens would often be beyond our reach, because we lack full empirical information about the consequences and circumstances of what (p. 127) the agent does, and so cannot know all the normatively significant act tokens she might perform. These shortfalls of knowledge are inevitable for any account of alternatives. But on **Highest**, merely ascertaining what the agent's alternatives *are* must remain forever beyond that agent's and our reach, even if it is known what acts the agent can physically perform and what normatively significant act types we should be on the lookout for.

4.3. Proposal III: Act Trees as Alternatives (Trees)

These considerations suggest that we should reject both **Bottom-most** and **Highest**, and explore the acceptability of **Trees** for defining alternatives. **Trees** says that an agent's alternative is an entire performable act tree, not any individual act token that constitutes part of that tree.[33]

Act trees are constituted by act tokens and their generational relations to each other. Two act trees are distinct from each other if they differ by at least one act token or one generational relationship. **Trees**, which identifies an agent's alternatives with the act trees that she might perform, has a number of virtues. First, it meets Constraint A, which stipulates that an entity counts as an alternative for an agent only if that entity is a concrete act and is performable by the agent. Act trees are not act tokens; rather they are comprised of act tokens. But act trees are not generic acts or act types, which Constraint A aims to exclude. Rather, act trees are structures of concrete entities, occurring at a particular time and performed by a particular agent, in the manner envisioned by Constraint A in focusing on act tokens. And **Trees** stipulates that an act tree must be performable if it is to count as an alternative, so it also meets this part of Constraint A.

Obviously **Trees**, which identifies an agent's alternatives as her possible act trees, avoids the first problem besetting **Bottom-most** and **Highest**, neither of which could appropriately define the act token putatively serving as an alternative without covertly invoking the agent's possible act trees as the real alternatives.

Furthermore, although at the bottom of two possible trees there may be act tokens that are agent-, time-, and act-type identical (such as moving one's finger), nonetheless some higher-level acts (such as donating $500 to Advance or donating $1,000 to Advance) or generational relationships on these trees will necessarily differ from each other since the trees themselves are distinct. So **Trees** evades the second problem afflicting Proposals **Bottom-most** and **Highest**, and it does not require problematic relabeling to avoid merging entities that we intuitively consider to be distinct alternatives. For this reason, if we understand an agent's set of alternatives (performable at a given time) to be all the act trees that the agent could perform at that time, the proposal appears to satisfy Constraint A's requirement that the acts in an alternative set are exhaustive of the agent's possibilities, and that they are mutually exclusive.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(p. 128) **Trees** also avoids the third problem of practicality encountered by **Highest**. According to **Highest**, we cannot identify which act qualifies as an agent's alternative unless we know all the normatively significant act types. But on **Trees** we can identify each of an agent's set of alternative act trees without having to know the act types of all the tokens that are members of those trees. Once an alternative is identified, we can set to work gaining more empirical and normative information about it. The severe practical difficulty encountered by **Highest** does not cripple **Trees**.

Trees should also meet Constraint B, which requires that an account of an agent's alternative set must be normative-theory neutral: it must identify the same set of actions as the agent's alternative set in the context of each plausible normative theory. Since an act tree contains all the act tokens that the agent would perform if she performed the basic acts on the tree in her circumstances, it perforce contains all the acts recognized as normatively significant by at least one plausible normative theory. No possible alternative that would be evaluated by some plausible normative theory will be left out of the alternative set.[34]

It appears, then, that we have found in **Trees** a satisfactory account of an agent's alternatives that fulfills both of the constraints we previously set forth and also avoids the problems afflicting both **Bottom-most** and **Highest**. It also appears to provide the foundation for addressing the questions about co-temporal acts and sequences of acts that have exercised theorists trying to resolve the question of how to identify an agent's alternatives. However, building on this foundation must be deferred to another occasion.

5. Relation between the Normative Status of an Act Tree and the Normative Status of Its Component Acts

I have now sketched and argued for **Trees**, which identifies an agent's alternatives with the set of act trees performable by her at a given time. This sketch leaves open the questions of how the normative status of an agent's alternative (an act tree) is related to the normative status of the act tokens that comprise that tree, and how the normative properties of an individual act token are related to those properties of other tokens on the same tree. Answering these questions is key to answering how **Trees** handles some of the normative issues raised by the existence of multiple "act versions." This section explores potential answers to these questions.

It is natural to think that whatever entity serves as one of the agent's alternatives must be the primary bearer of its normative properties: normally we judge *alternatives* as (p. 129) right or obligatory, or forbid them as wrong. At the same time, we also normally characterize act tokens, such as breaking a promise, as being right or wrong. Given **Tree**'s rejection of tokens as alternatives, can we reconcile these intuitions? Satisfying

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

these intuitions should comprise our first two criteria for acceptability of any proposal regarding how the normative properties of tokens and trees are related.

5.1. Possible Relations between Normative Properties of Act Trees and Those of Their Component Act Tokens

There are undoubtedly multiple possible ways to resolve the question of this relationship. I shall briefly describe five options as leading candidates and then point out some of the advantages and disadvantages of each.

(Option1) ***Trees alone***: Act trees are the sole bearers of normative properties (e.g., *being pro tanto wrong*), in virtue of the act types of their component tokens. In the case of comparative normative properties (e.g., *being ATC wrong*), a tree's possession of such a property depends both on the act types of its component act tokens and on the act types of the component act tokens of alternative trees. Tokens themselves have no normative properties, even though they may be tokens of normative act types (*doing what is pro tanto wrong)*, or, as in the case of a token such as breaking a promise, may generate tokens of normative act types.

(Option 2) ***Trees first***: Act trees are the primary and only direct bearers of normative properties, in virtue of the normative act types of their component tokens (and in the case of comparative normative properties, also in virtue of the normative act types of the component tokens of alternative trees). Tokens can inherit normative properties from the trees of which they are components. In such cases, the tokens have the normative properties indirectly.

(Option 3) ***Tokens alone***: Only act tokens have normative properties. On one version of ***Tokens alone***, the only tokens bearing normative properties are those that have them directly in virtue of their act types. On another version, tokens may bear normative properties directly in virtue of their act types, or they may inherit normative properties indirectly in virtue of their relations to tokens on the same tree that directly bear normative properties.

(Option 4) ***Tokens first***: Act tokens are the primary bearers of normative properties, in virtue of their act types. Trees (and other tokens on the same tree) can inherit these properties from such tokens.

(Option 5) ***Trees and tokens together***: Both act tokens and act trees have normative properties both directly and indirectly.

Each of these options has its strengths and weaknesses, which there is insufficient room here to explore thoroughly. Let us examine each of them briefly.

(p. 130) (Option 1) *Trees alone* holds that act trees are the sole bearers of normative properties; act tokens have no such properties. This view has the virtue that it ascribes normative properties to an agent's alternatives, namely her act trees. However, it has the disadvantage that it declines to ascribe normative properties to any of an agent's act tokens, thus flying in the face of our normal practice of evaluating act tokens, such as Lisa's breaking her promise, as morally wrong. ***Trees alone*** fails to meet our second criterion.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Trees alone has a serious structural flaw as well. In introducing conventional generation, I said that the token breaking a promise conventionally generates the token of doing something pro tanto wrong in circumstances in which there is a rule that breaking promises is pro tanto wrong. Implicitly this ascribes the normative property of *being pro tanto wrong* to the act of breaking a promise. Similarly, Rosa's act of shooting the gun has the property of *causing Henry's death*, and so generates, by causal generation, Rosa's act of killing Henry. But ***Trees alone*** does not allow tokens to have normative properties, so the token of breaking promises has no such property and so cannot generate a token of doing something pro tanto wrong. This has devastating implications. First, contrary to our normal assumption, there are no act tokens of *doing something pro tanto wrong* or any other normative act type. Second, ***Trees alone*** holds that act trees, the sole bearers of normative properties, have those properties in virtue of the act types of their component act tokens. But since the tokens have no normative act types, their trees cannot have their normative properties in virtue of such act types (or in virtue of their tokens having normative properties). Nor does there seem to be any obvious alternative way for trees to acquire normative properties. A supporter of ***Trees alone*** might claim that occurrence of an act token of the type *breaking a promise* directly imbues its tree with the property of *being pro tanto wrong*. But now we need an explanation of why this is true, when it's false that this act token has a property that enables it to generate a *token* of the type *doing something pro tanto wrong*. The views embodied in ***Trees alone*** seem either arbitrary or subject to crippling inconsistency.

(Option 2) ***Trees first*** holds that act trees are the primary and only direct bearers of normative properties, in virtue of the normative act types of their component tokens. Tokens can inherit normative properties from the trees of which they are components; the tokens then bear these properties indirectly. This view fulfills the two key requirements that both tokens and trees have normative properties, and that trees are the primary bearers of such properties.

But on this view, in virtue of what does an act tree possess a normative property, such as *being pro tanto wrong*? One tempting answer is that one of the tree's tokens has this property, and the tree inherits it from the token. But this would contravene a central tenet of ***Trees first***, namely that all normative properties originate as the properties of trees. A second possible answer is that one of the tree's tokens is of the type *doing something pro tanto wrong*, and the tree acquires the normative property of *being pro tanto wrong* from this token. But as we've just seen in assessing ***Trees alone***, a token is only of the type *doing something pro tanto wrong* if it is generated by a lower act, such as breaking a promise, which has the normative property *being pro tanto wrong*. However, ***Trees first*** precludes any token from having a normative property except via inheritance from (p. 131) its tree. ***Trees first***, then, cannot explain how act trees (and ultimately act tokens) come to have normative properties, and it denies what seems obvious, namely that act tokens have their normative properties in virtue of their act types. ***Trees first***, then, does not seem to be a satisfactory proposal for how the normative properties of trees and tokens are related to each other.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(Option 3) ***Tokens alone*** holds that only tokens have normative properties, either directly in virtue of their act types or indirectly by inheritance from another act token on the same tree. It satisfies the requirement that act tokens possess normative properties, and (unlike ***Trees alone*** and ***Trees first***) it provides a satisfactory explanation for how this comes about. However, since it denies that act trees—the agent's alternatives—have normative properties, it fails to honor our key assumption that alternatives are the primary bearers of normative properties. This proposal should be rejected as not meeting our first criterion.

(Option 4) ***Tokens first*** holds that tokens are the primary bearers of normative properties, in virtue of their act types. Trees (and other tokens on the same tree) can inherit these properties from such tokens.

Tokens first satisfies our two key criteria, since it ascribes normative properties to both act tokens and act trees (the agent's alternatives). It may not accord *primary* status to trees for any normative property, but this may not be a fatal problem.

However, ***Tokens first*** may not be sufficiently nuanced. As stated, it suggests that any normative property of a token is inherited by the token's tree. For many properties this seems perfectly appropriate. If Lisa's act of breaking her promise is pro tanto wrong, then the act tree of which it is a component is pro tanto wrong as well. Another token on this same tree, her saving five lives, is pro tanto right, so the act tree is pro tanto right—as well as pro tanto wrong. As suggested in section 3, other cases may be less clear, including cases of trees inheriting the properties of being blameworthy and being supererogatory from their act tokens. If only certain normative properties—call them "core" normative properties—are inherited by trees from their tokens, then ***Tokens first*** needs to be stated in a more nuanced fashion to capture this fact. Dealing with these nuances would require work, but this does not seem to be a serious problem for ***Tokens first.***

A more complex issue is raised by essentially comparative normative properties, such as *being ATC wrong*. ***Tokens first*** assumes that the normative properties of act tokens are the originating source of an act tree's normative properties. However, this is difficult to maintain for essentially comparative normative properties. A comparative normative property characterizes an entity in virtue of a comparison to its alternatives. Thus an act token can only have the property of *being ATC wrong* if it is morally worse than at least one of its alternatives. But according to **TREES** (our account of alternatives), an act token is not an alternative, and so nothing qualifies as an alternative to it. No token, then, can directly have an essentially comparative normative property. According to **TREES**, act trees are alternatives to each other. According to ***Tokens first***, trees can only acquire normative properties by inheritance from their tokens. But no tree can inherit a comparative normative property from its tokens, which lack such properties. Thus according to ***Tokens first***, neither tokens nor trees can possess (p. 132) (directly or indirectly) essentially comparative normative properties. This deficiency rules out ***Tokens first*** as unacceptable.[35]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(Option 5) This leaves ***Trees and tokens together***, which holds that both act tokens and act trees have normative properties both directly and indirectly. Although ***Trees and tokens together*** (hereafter, ***T&TT***) is the most complex proposal, on the whole it seems to me the most promising, in terms of capturing the key normative judgments about actions that we routinely make, and in terms of avoiding possible problems. A primary virtue of ***T&TT*** is that it satisfies our two intuitions that (1) an agent's alternatives (that is, act trees) are bearers (perhaps the primary bearers) of the normative properties of her conduct, and (2) act tokens themselves have moral (and other normative) properties. One natural way to develop the essential features of ***T&TT*** would involve making three stipulations. First we stipulate that an act token can directly have a core moral property (such as *being pro tanto wrong*) in virtue of its act type (so Lisa's act of breaking her promise has the moral property of *being pro tanto wrong*). Second we stipulate that an act tree inherits core noncomparative moral properties from its component tokens in virtue of their directly held moral properties. For example, Lisa's act tree ***A***, which involves giving $1,000 to Advance, inherits the moral property of *being pro tanto wrong*, because one of its component act tokens (her breaking a promise) directly has that moral property. Note this implies that core noncomparative normative properties are held in the first instance by act tokens, and only indirectly by act trees.

The third stipulation regarding ***T&TT*** is that an act tree directly has comparative moral properties, such as *being ATC wrong*, in virtue of both the moral properties of its component act tokens *and* the moral properties of component act tokens of alternative act trees.[36] For example, Lisa's tree ***A*** has the moral property of *being ATC wrong*, since (a) several of its act tokens have the moral property of *being pro tanto wrong*, (b) several of its tokens have the moral property of *being pro tanto right*, (b) Lisa's alternative act tree ***B*** (which involves her giving $1,000 to CARE rather than Advance) has a more favorable balance than act tree A of act tokens having the properties of *being pro tanto right* and *being pro tanto wrong*, and (d) in virtue of these facts, her act tree ***A*** has the property of *fulfilling a smaller net balance of pro tanto duties than some alternative act tree* and hence the comparative moral property of *being ATC wrong*. The act tree is the originating source of this latter property, which its act tokens can inherit. This implies that comparative normative properties are held in the first instance by act trees, and only indirectly by act tokens. This reveals a more subtle version of our original intuition that agent's alternative must be the *primary*—that is, direct—bearer of its normative properties. Act trees (p. 133) are the agent's alternatives; but they are the primary bearers of essentially comparative normative properties, not of all normative properties.

Exactly which directly held normative properties of act tokens are inherited by their trees, and exactly which directly held normative properties of act trees are inherited by their component tokens, is a matter of complex judgment (as we saw in discussing ***Tokens first***). Further work is required to settle this as well as to settle which normative properties a token can inherit from other tokens on the same tree. However, resolving these issues is a matter of developing ***T&TT***, not a matter for wholesale revision. Deciding some of these questions, such as how to treat blameworthiness and supererogatori-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

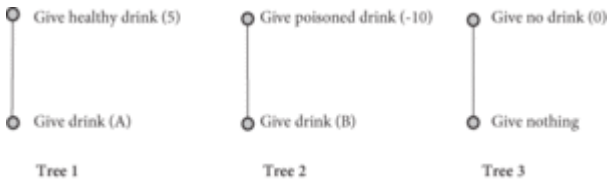
ness, will not necessarily be easy. However, if we accept **TREES**, we will not have to decide them in order to determine which entities to recognize as an agent's alternatives.

What issues does ***T&TT*** enable us to avoid? Unlike ***Tokens alone*** and ***Trees alone***, it satisfies both our two key intuitions that normative properties are held by both tokens and trees. Also unlike ***Trees alone*** and ***Trees first***, ***T&TT*** provides a natural explanation for why trees have the normative properties they have either directly or indirectly, including comparative properties. And it accords with our intuition that an act token has a noncomparative act property in virtue of its act type. Unlike ***Tokens first***, ***T&TT*** gives primary status to trees as the originating source for essentially comparative normative properties such as *being ATC wrong*. Among our five options, then, ***T&TT*** seems to provide the best account of how the normative properties of tokens and trees are related to each other.

5.2. Addressing the Traditional Problem Created by Multiple Act Versions

The traditional problem of act versions arises because there are multiple "versions" of an act. We have recharacterized these multiple versions as multiple act tokens on the same act tree. Such versions can have different causal consequences from each other and, assuming certain plausible principles of deontic logic, can generate inconsistent normative prescriptions.

An excellent description of the inconsistency problem is outlined by Timmerman and Cohen (2019, section 6), based on presentations by Brown (2018) and Portmore (2017). Consider the view of "Omnists," who hold that all acts should be directly assessed in terms of their possession (or not) of right-making properties. Imagine an agent Sam, who has the option of giving the king a healthy drink, giving the king a poisoned drink, or giving the king nothing to drink.[37] Giving the king a healthy drink would produce 5 hedons, giving him a poisoned drink would produce –10 hedons, and giving him nothing to drink would produce 0 hedons. Using our concepts, we can see that giving the king a healthy drink would be generated by giving him a drink, giving him a poisoned (p. 134) drink would be generated by giving him a drink, and giving him nothing to drink would be generated by giving him nothing. We can represent this by Figure 6.2, a diagram of three possible act trees.



(p. 136) *Figure 6.2.* Sam's alternatives

Omnism is held to imply, assuming hedonistic consequentialism, that Sam's giving the king a healthy drink is obligatory, since this act token has the right-making property of

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

producing more hedons than either giving him a poisoned drink or giving him no drink. But suppose it is true that if Sam gave the king a drink, he would give him a poisoned drink. Then giving the king a drink would produce fewer hedons (–10) than giving him nothing (0 hedons), so giving him a drink would be wrong. Omnists assume the truth of the following commonly accepted deontic principle, the Principle of Normative Inheritance:

(NI): If act X entails act Y, then X's being obligatory entails Y's being obligatory.

Discussants of Omnism and its rivals assume that (for example) giving the king a healthy drink "entails" giving him a drink, either because there is some logical or semantic entailment relation between these two acts, or because (more plausibly) it's not metaphysically or practically possible to give the king a healthy drink without giving him a drink. However, Omnists have now made a set of assumptions which jointly entail the fatally inconsistent conclusion that giving the king a drink is both wrong (because it would produce fewer good consequences than giving him nothing), *and also* obligatory (because it is entailed by the obligatory act of giving him a healthy drink).[38]

How would ***Trees and tokens together*** (***T&TT***) handle Sam's case? Let's continue to evaluate Sam's acts from the perspective of hedonic consequentialism. Standard hedonic consequentialism only countenances essentially comparative moral properties, such as *all-things-considered wrongness*. On ***T&TT*** no act token directly has such a comparative moral property, since the comparison required is among act trees. Thus a tree such as Tree 2 directly possesses such a comparative moral property in virtue of the nonmoral properties of its own act tokens (such as *producing –10 hedons*) and the nonmoral properties of the act tokens of its alternatives, Trees 1 and 3. Tree 2 has the essentially comparative moral property of *being ATC wrong*. On a natural version of ***T&TT***, all the (p. 135) tokens of Tree 2 inherit this moral property. So giving the king a poisoned drink, and giving him a drink (B) that generates it, are both indirectly ATC wrong. By similar reasoning, Tree 1 is ATC obligatory, so its tokens giving the king a healthy drink and giving him a drink (A) are both indirectly ATC obligatory. As we can see, in this case ***T&TT*** avoids the second traditional problem of act versions, since it does not produce any inconsistent evaluations of a single act entity, either an act tree or an act token. Nor is there reason to fear inconsistency would arise in the context of other moral theories.[39]

What about the claim that if Sam gave the king a drink, he would give him a poisoned drink? On its face, this claim has no determinate truth value, since Sam's giving the king a drink could refer either to token (A) or token (B), only one of which generates giving the king a poisoned drink.[40] For this claim to have a determinate truth value, it would have to be interpreted as a more complex counterfactual, such as "If Sam were to perform an act of the type *giving the king a drink*, he would do so in a way that would generate his poisoning the king."[41] Suppose this complex counterfactual is true because Sam wants to poison the king. Then he genuinely has the three alternative act trees depicted in Figure 6.2, of which he would perform Tree 2. According to ***T&TT***, this implies that giving-the-king-a-drink$_B$ is ATC wrong, while giving-the-king-a-drink$_A$ is ATC obligatory. The moral

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

status of each of these acts is inherited from its act tree. Even if token (A) were comparatively evaluated in terms of its own consequences (not genuinely possible according to **TREES**, since token (A) is not an alternative and itself has no defined alternatives), those consequences would be better than those of token (B) or of Sam's giving the king nothing to drink. So token (A) could not be evaluated as ATC wrong. No inconsistency arises: token (A) is not both obligatory and wrong. No evaluation can be assigned to "the act" of giving a drink to the king, since there is no single act of this type. Any evaluation must be of token (A) or token (B), each of which inherits its moral status from its own tree.

Of course, we can still ask what Sam ought to do, given that he's not going to perform Tree 1. This is a substantive normative question. If we pick the right entities to serve as alternatives, we won't be led down the wrong path in seeking an answer to it.[42]

6. Conclusion

For several decades theorists have debated how to define an agent's alternatives for purposes of normative theory, focusing on questions that arise because (a) at any given time an agent may perform numerous "versions" of actions, (b) an agent may be able to perform several independent but simultaneous co-temporal actions, and (c) agents have the ability to perform whole sequences of actions in which performance of one act in a sequence may affect the performance or normative status of others in the sequence. To answer these questions, I have posited that we need to equip ourselves with a robust theory of human action enabling us to understand the issues more clearly. After outlining Alvin Goldman's fine-grained account of action, I have employed it to address the problems raised by question (a), arguing that we should accept **TREES**, according to which an agent's alternative at a time is identified as an entire act tree performable by her, rather than as any of the act tokens that comprise the act tree. I have further employed this action theory in arguing that ***Trees and tokens together*** is the most promising account of how the core moral evaluations of an act tree and its component act (p. 137) tokens are related. These proposals lay the groundwork for future work addressing questions (b) and (c).[43]

References

Anscombe, G. E. M. 1958. *Intention.* Ithaca, NY: Cornell University Press.

Åqvist, Lennart. 1969. "Improved Formulations of Act-Utilitarianism." *Noûs* 3: 299–323.

Archer, Alfred. 2016. "Moral Obligation, Self-Interest and the Transitivity Problem." *Utilitas* 28: 441–464.

Benn, Claire. 2018. "Supererogation, Optionality, and Cost." *Philosophical Studies* 175: 2399–2417.

Bennett, Jonathan. 1973. "Shooting, Killing, and Dying." *Canadian Journal of Philosophy* 2: 315–322.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Bergström, Lars. 1966. *The Alternatives and Consequences of Actions*. Stockholm: Almqvist & Wiksell.

Bergström, Lars. 1971. “Utilitarianism and Alternative Actions.” *Nôus* 5: 237–252.

Brown, Campbell. 2018. “Maximalism and the Structure of Acts.” *Nôus* 52: 752–771. doi: 10.1111/nous.12181.

Bykvist, Krister. 2002. “Alternative Actions and the Spirit of Consequentialism.” *Philosophical Studies* 107, no. 1: 45–68.

Carlson, Erik. 1999. “Consequentialism, Alternatives, and Actualism.” *Philosophical Studies* 96, no. 3: 253–268.

Castañeda, Hector-Neri. 1968. “A Problem for Utilitarianism.” *Analysis* 28: 141–142.

Davidson, Donald. 1963. “Actions, Reasons, and Causes.” *The Journal of Philosophy* 60, no. 23: 685–700.

Feldman, Fred. 1975. “World Utilitarianism.” In *Analysis and Metaphysics,* Philosophical Studies Series in Philosophy 4, edited by Keith Lehrer, 255–271. Dordrecht, the Netherlands: Springer.

Glasgow, Joshua. 2012. “Kant’s Principle of Universal Law.” In *Conduct and Character; Readings in Moral Theory*, 6th ed., edited by Mark Timmons, 152–165. Boston: Wadsworth.

Goldman, Alvin. 1970. *A Theory of Human Knowledge*. Englewood Cliffs, NJ: Prentice-Hall. (Reprinted by Princeton University Press, 1977).

Goldman, Holly Smith. 1976. “Dated Rightness and Moral Imperfection.” *The Philosophical Review* 85, no. 4: 449–487.

Goldman, Holly Smith. 1978. “Doing the Best One Can.” In *Values and Morals*, edited by Alvin Goldman and Jaegwon Kim, 186–214. Dordrecht, the Netherlands: Reidel.

Gustafsson, Johan. 2014. “Combinative Consequentialism and the Problem of Act Versions.” *Philosophical* Studies 167, no. 3: 585–596.

Jackson, Frank, and Pargetter, Robert. 1986. “Oughts, Options, and Actualism.” *The Philosophical Review* 95: 233–255.

Kiesewetter, Benjamin. 2015. “Instrumental Normativity: In Defense of the Transmission Principle.” *Ethics* 125: 921–946.

(p. 138) Levine, Sydney, Leslie, Alan M., and Mikhail, John. 2018. “The Mental Representation of Human Action.” *Cognitive Science* 42, no. 4: 1229–1264.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Lombard, Lawrence B. 1978. “Actions, Results, and the Time of a Killing.” *Philosophia* 8, no. 2–3: 341–354.

Mikhail, John. 2011. *Elements of Moral Cognition: Rawls’ Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. Cambridge: Cambridge University Press.

Nell, Onora. 1975. *Acting on Principle*. New York: Columbia University Press.

Portmore, Douglas. 2017. “Maximalism versus Omnism about Permissibility.” *Pacific Philosophical Quarterly* 98(S1): 427–452.

Prawitz, Dag. 1968. “A Discussion Note on Utilitarianism.” *Theoria* 34: 76–84.

Prawitz, Dag. 1970. “The Alternatives to an Action.” *Theoria* 36: 116–126.

Smith, Holly M. 2018. *Making Morality Work*. Oxford: Oxford University Press.

Sobel, Jordan Howard. 1976. “Utilitarianism and Past and Future Mistakes.” *Noûs* 10: 195–219.

Thomson, Judith Jarvis. 1971. “The Time of a Killing.” *The Journal of Philosophy* 68, no. 5: 115–132.

Timmerman, Travis, and Cohen, Yishai. 2019. “Actualism and Possibilism in Ethics.” In *The Stanford Encyclopedia of Philosophy* (Summer 2019 edition), edited by Edward N. Zalta. **https://plato.stanford.edu/archives/sum2019/entries/actualism-possibilism-ethics/**.

Weintraub, Ruth. 2003. “The Time of a Killing.” *Analysis* 63, no. 279: 178–182.

Notes:

([1]) See, for example, Bergström (1971); Gustafsson (2014); and Portmore (2017).

([2]) See Bergström (1966); Prawitz (1968); Bykvist (2002); Gustafsson (2014); Brown (2017); and Portmore (2017).

([3]) See, for example, Prawitz (1968); Carlson (1999); Bykvist (2002); Brown (2017); and Portmore (2017).

([4]) See, for example, Prawitz (1968); Carlson (1999); Bykvist (2002); Brown (2017); and Portmore (2017). This is a variant of a case from Timmerman and Cohen (2019).

([5]) Following A. Goldman (1970), 22.

([6]) See Prawitz (1970); Feldman (1975); H. S. Goldman (1978); Carlson (1999); Gustafsson (2014); Brown (2017); and Portmore (2017).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([7]) See Prawitz (1970); Åqvist (1969); H. S. Goldman (1976, 1978); Sobel (1976); Jackson and Pargetter (1986); Carlson (1999); Bykvist (2002); Gustafsson (2014); Archer (2016); Brown (2017); and Portmore (2017).

([8]) At least for the kinds of actions I discuss. Both Douglas Portmore and James Goodrich (private communications) have pointed out that there are mental as well as bodily acts. These may require a different analysis, since the individuation issues may be different.

([9]) Brown (2017) raises but doesn't try to answer the question whether it is a problem for, say, Kantianism. Portmore (2017) considers cases in which an option that is optimal in terms of consequentialist factors is also optimal in terms of nonconsequentialist factors such as justice, fidelity, respect for autonomy, and so forth.

([10]) Glasgow's discussion is based on Nell (1975).

([11]) What to say about the time interval of an act is a highly contested issue. Charlene's firing the gun may happen at t_1. But suppose the victim dies from his wound a month later, at t_{30}. What is the time interval of Charlene's act of killing the victim? For purposes of this paper I shall not delve more deeply into this issue. See Thomson (1971); Bennett (1973); Lombard (1978); and Weintraub (2003).

([12]) These conditions originate with Bergström (1966).

([13]) As Caspar Hare points out (personal communication), subjectivist moral theories may utilize different alternatives than objectivist moral theories. If so, Constraint B would require modification.

([14]) As James Goodrich notes (personal correspondence), there might be incomplete but sufficient overlap in the alternatives recognized by two theories, at least in particular cases, to allow comparisons between the acts they prescribe. However, it seems reasonable to want a guarantee of adequate overlap for all theories that might be compared, and Constraint B is the most straightforward way of securing this guarantee.

([15]) Many theorists working on the problem of defining alternatives seem to have implicitly accepted some form of a fine-grained approach to act individuation. See Prawitz (1968) and Bykvist (2002). Mikhail (2011, chap. 6) and Levine et. al. (2018) have made innovative use of Goldman's action theory in the law and in psychology. Because the fine-grained approach, as developed by Goldman, highlights the structural relations among act tokens, it is especially well-adapted for use in addressing our problem. How this might be done with a Davidsonian approach remains unclear to me.

([16]) The following account is summarized from A. Goldman (1970), chaps. 1, 2, and 3. Act types are sometimes called "generic acts." See Bergström (1971).

([17]) For brevity I shall not try to define what it is for an act token to be performable. Discussing this would lead us into questions of free will, which are certainly germane to the question of defining an agent's alternatives, but they are not the focus of this chapter.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([18]) For Goldman's detailed discussion of basic acts, see A. Goldman (1970), Chapter 3, section 4. Contemporary work in metaphysics might classify generation as a species of grounding.

([19]) Several theorists in discussing alternative actions argue that we must restrict our attention to intentional acts, unlike Lisa's violating her promise (Carlson 1999; Bykvist 2002). This seems a mistake, since (on most moral theories) nonintentional actions, such as Lisa's violating her promise, have moral status (such as being wrong but not blameworthy), just as intentional actions do.

([20]) Of course, although the proposition "Lisa moves her finger slowly" entails the proposition "Lisa moves her finger," the first *act* does not literally *entail* the second act, as this wording would suggest, since acts are events rather than linguistic entities that can stand in entailment relations. In discussing act versions some theorists speak about acts entailing other acts; other theorists say things like "The occurrence of act ***a**** necessitates the occurrence of act ***a***." For the latter usage, see, for example, Brown (2017) and Portmore (2017).

([21]) Brown (2017, 1, 9) apparently would claim (contrary to Goldman) that Lisa's moving her finger was done by her moving her finger slowly. He is correct, however, in saying that Lisa's moving her finger slowly "entails" her moving her finger.

([22]) A. Goldman (1970), 33. Although Goldman standardly views act trees as diagrams, I shall use the term to refer both to a certain metaphysical structure, comprising act tokens and their relationships, and to a diagram illustrating that structure.

([23]) This feature is a key part of the argument in favor of the fine-grained view as opposed to the coarse-grained view, since X is identical with Y only if X and Y have all their properties in common. See A. Goldman (1970), chap. 1.

([24]) *Blameworthiness* and *praiseworthiness* are clearly moral properties, but they do not belong to the same range of deontic properties as do *obligatory*, *wrong*, and *right*. However, since on some views certain acts are wrong only if they are performed intentionally, on these views even the property of *being wrong* is one that, like *being blameworthy*, could in some cases only apply to an intentional act.

([25]) See Benn (2018), Section 2.1. Note that act tokens can have properties (such as occurring on Monday) that do not arise in virtue of the token's act type.

([26]) See, for example, Brown (2017), 3–4; Portmore (2017), 429; and Kiesewetter (2015). Kiesewetter lists other supporters and opponents of "inheritance" or "transmission" principles.

([27]) This formulation should be extended to include what Goldman terms "same level acts," such as checkmating one's opponent and checkmating Bobby Fisher, when Bobby Fisher is one's opponent (Goldman 1970, 30–31).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([28]) Note, as Doug Portmore has pointed out to me, that this proposal apparently can't address what counts as an alternative when one option is a mental act, such as thinking of a pink elephant.

([29]) Note we are discussing act tokens in different possible worlds.

([30]) Lisa's giving to charity generates her donating to Advance by augmentation generation.

([31]) For a discussion of the Usability Demand for normative theories, see Smith (2018).

([32]) Of course it would not be impossible if ***a*** is the lowest act token. But then one has fallen back on **Bottom-most.**

([33]) This proposal would require further development to extend to co-temporal acts and sequences of acts.

([34]) Of course, Constraint B is not a constraint on what metaphysical structures can count as act trees. Rather it is a constraint on what kinds of entities can count as alternatives.

([35]) A similar problem afflicts ***Tokens alone.***

Hallie Liberto (in discussion) suggests that, under **Tokens first**, trees could acquire essentially comparative normative properties (such as *being ATC obligatory*) in virtue of the various nonnormative properties (such as *producing N amount of happiness*) possessed by a tree and its alternatives. However, this is barred by the stipulation in **Tokens first** that any normative property must be initially possessed by a token, and only be inherited by its tree.

([36]) Alternatively, in virtue of the noncomparative nonnormative properties it and its alternative trees inherit from their tokens.

([37]) This is a variant of Timmerman and Cohen's example.

([38]) Note that this argument assumes that the tokens giving-a-drink$_A$ and giving-a-drink$_B$ are identical to each other. Because these two tokens are agent-, time- and type-identical, it's tempting to fall into this assumption. However, since they occur on distinct act trees, they are actually distinct tokens.

([39]) Given the restriction of standard consequentialism to essentially comparative moral properties, within the context of this moral theory, no act token has a core moral property directly. A deontological theory would ascribe noncomparative moral properties, such as *being pro tanto wrong*, directly to act tokens. But there is no inconsistency in a tree being ATC obligatory while one of its tokens is (for example) pro tanto wrong.

([40]) A parallel would be the indeterminate truth value of "If the governor signed a bill, it would be a tax-cut bill," when it is indeterminate which governor is being referred to.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Note this is a place where thinking of “giving the king a drink” as a single act token can mislead us.

([41]) I owe this suggestion to Niko Kolodny. Another possible interpretation is “As between giving the king a poisoned drink and giving him a healthy drink, Sam would give him a poisoned drink.”

([42]) At this point we can see another arena in which the fine-grained action theory can help us see certain issues more clearly. In discussing the problem of act versions, Portmore states that “the performance of one option can entail the performance of another. For instance, I have the option of baking a pie as well as the option of baking, and baking a pie entails baking” (Portmore 2017, 427). He explains his notion of “entailment” as follows: “[F]or any two options φ and ψ, φ-ing entails ψ-ing if and only if φ-ing without ψ-ing isn’t an option” (Portmore 2017, 427). He then introduces the concept of a “maximal” action by stating that “A maximal option is an option that is maximally normatively specific in the sense that it is entailed only by normatively equivalent options (which, of course, includes itself)” (Portmore 2017, 428). He continues by defining his preferred position, Maximalism, as the view that, when *p* is a right-making property of actions, (1) a maximal option which has *p* is right, but (2) a nonmaximal option, whether or not it has *p*, is right if and only if it is entailed by a maximal option (Portmore 2017, 429).

Clearly baking a pie “entails” baking: one can’t bake a pie without baking. But what would Portmore say about a case in which the agent has a choice between two ways to do something? For example, Lisa can donate to charity by moving her finger in way X or moving her finger in way Y. Let’s suppose her donating to charity is a maximal act, and one that is obligatory. Suppose she actually moves her finger in way X. Then, on Portmore’s view, does her donating to charity entail her moving her finger in way X? Interpreting his definition of “entail” one way, the answer is “no,” since she could donate to charity by moving her finger in way Y rather than by moving her finger in way X. But interpreted in another way, the answer is “yes,” since her donating to charity *in the way she does—that is, the act token she actually performs*—could only happen if she moves her finger in way X. Suppose Portmore opts for the first interpretation. Then Lisa’s moving her finger in way X is a nonmaximal act, and since it is not entailed by her donating to charity, it is neither obligatory, nor right, nor wrong. It inherits no moral property. This seems strange, since it is only by moving her finger in way X that she does something obligatory. Her moving her finger is her means of donating to charity, and we would normally assume the means inherits the moral property of the act for which it is a means. Suppose instead Portmore opts for the second interpretation, according to which her moving her finger in way X is entailed by her donating to charity. Then her nonmaximal act of moving her finger in way X is morally obligatory. This seems like the more plausible assessment. Moreover, when we understand that the entailment relations are between act tokens, the second interpretation seems like the only feasible one.

But on this interpretation it appears that it would be more illuminating to understand the relationship between entailing and entailed acts as the relationship between generated

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

and generating acts, since Lisa’s moving her finger in way X generates her donating to charity. Thinking about such cases with the help of action theory, the generation relationship, and identity conditions for act tokens allows us to discern certain issues that otherwise might have remained invisible and to better decide what to say about them.

([43]) I would like to thank Alvin Goldman, Niko Kolodny, Douglas Portmore, the audiences at the 2019 Chambers Conference at the University of Nebraska-Lincoln and the Practical Philosophy Seminar at the University of Stockholm, and especially James Goodrich and Nikolaj Nottelman, for helpful comments on earlier versions of this material.

Holly M. Smith

Holly M. Smith is Distinguished Professor Emerita of Philosophy at Rutgers University and Distinguished Research Associate at The University of California, Berkeley. She has also held appointments at Tufts University, the University of Pittsburgh, the University of Michigan, the University of Illinois-Chicago, and the University of Arizona. Her publications principally focus on topics in normative ethics, moral decision making, the theory of moral responsibility, and biomedical ethics. In Making Morality Work (Oxford University Press, 2018), she explores how moral theories should accommodate the errors, ignorance, and misunderstandings that impede us as moral decision makers. Her current projects propose new strategies for weighing the stringency of deontological duties, and for identifying and evaluating an agent’s alternatives in the context of normative theories.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Value Comparability

Alastair Norcross

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.3

Abstract and Keywords

Value comparability, either in the form of the kind of quantitative comparability involved in (intrapersonal or interpersonal) aggregation, or in the form of the kind of qualitative comparability involved in comparing putatively different values, has been thought to threaten the theoretical soundness of consequentialist theories. In part 2, I argue that unrestricted axiological aggregation is supported by overwhelmingly plausible assumptions about ordinary value comparisons. In particular, I argue that large numbers of small harms, such as headaches, really can outweigh small numbers of large harms, such as deaths. In part 3, I consider the challenge that qualitatively different values may be incomparable, in the sense that instances of one value may be neither better, worse, nor equal in value with instances of a different value. I argue that all values, no matter how qualitatively distinct, are either thoroughgoingly comparable or not at all (and that the latter is too implausible to take seriously).

Keywords: aggregation, comparability, incommensurability, qualitative, quantitative, values

1. Introduction

CONSEQUENTIALIST theories traditionally consist in two parts: an axiological theory and a deontic theory. The axiological theory tells us what is intrinsically valuable, or to be promoted. The deontic theory tells us how our actions relate to what is to be promoted. The best-known form of consequentialism is maximizing act utilitarianism, which tells us (roughly) that actions are right, or obligatory, just in case they maximally promote the net[1] amount of happiness, or well-being, in the world; otherwise they are wrong.[2] This combines a welfarist axiology with a maximizing deontic theory. Not all versions of consequentialism incorporate a maximizing theory of right action,[3] or any theory of right action for that matter.[4] And different versions of consequentialism incorporate different theories of the good, different axiologies. But all share the view that the comparative amounts of good resulting from different choices generate moral reasons to prefer different choices. Thus, there is moral reason to prefer the choice leading to the greatest good over one

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

leading to less good, the strength of the reason being proportional to the size of the difference in good. This, of course, assumes that we can compare outcomes (or worlds) in terms of their overall value. Some challenges to consequentialism consist in challenges to this assumption, which we can call the assumption of "value comparability."

(p. 359) In this article, I will examine and respond to two different challenges to the kind of value comparability incorporated by pretty much any version of consequentialism. First, there is what we might think of as a challenge to *quantitative* value comparability: the claim that harms and benefits of significantly different sizes can be traded off against each other in order to justify imposing a small number of large harms on some to bring a large number of small benefits (such as the prevention of small harms) to some. Since the objection to these kinds of value comparisons is always put in terms of the imposition of large harms on some people[5] in order to prevent small harms to *other* people, it is helpful to think of this objection as a challenge to *interpersonal aggregation*. Second, there is the potentially more radical challenge to *qualitative* value comparability: the claim that we can compare the overall, or all-things-considered, values of total states of affairs containing values of radically different kinds, such as pleasure and integrity, or even, within a hedonic value theory, *pleasures* of radically different kinds, such as the pleasure of eating chocolate and the pleasure of emotional satisfaction. Philosophers who object to qualitative value comparability often put their challenge as the claim that (at least some) values are "incommensurable."

2. Challenging Interpersonal Aggregation

Henry Sidgwick (1981) suggests as the "maxim of Rational Self-Love or Prudence" that "one ought to aim at one's own good." The problem, he says, with this maxim is that it is tautological, since "one's own good" can be defined as "what one ought to aim at." However, if we modify the maxim to say that one ought to aim at one's own good *on the whole*, we avoid tautology and arrive at what he calls the principle of "impartial concern for all parts of our conscious life." This principle, he says, is applicable to "any … interpretation of 'one's own good' in which good is conceived as a mathematical whole, of which the integrant parts are realized in different parts or moments of a lifetime." So, it is part of prudence, for example, to forego a lesser present pleasure in exchange for a greater future pleasure (or absence of greater future pain). This requires a comparison of different goods within one's own lifetime. The more controversial commitment comes in the move from the prudential to the moral. Here is Sidgwick again, at greater length:

> So far we have only been considering the "Good on the Whole" of a single individual: but just as this notion is constructed by comparison and integration of the different "goods" that succeed one another in the series of our conscious states, so we have formed the notion of Universal Good by comparison and integration of the (p. 360) goods of all individual human—or sentient—existences. And here again, just as in the former case, by considering the relation of the integrant parts to the whole and to each other, I obtain the self-evident principle that the good of any

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> one individual is of no more importance, from the point of view (if I may say so) of the Universe, than the good of any other … From these two rational intuitions we may deduce, as a necessary inference, the maxim of Benevolence in an abstract form: viz. that each one is morally bound to regard the good of any other individual as much as his own, except in so far as he judges it to be less, when impartially viewed.(Sidgwick 1981, III, 13, 3)

In moving from prudence to benevolence, we have to extend our value comparisons beyond the intrapersonal to the interpersonal. If I am to regard the good of another as much as my own, except when I judge it to be less, I must be able to compare the good of another with my own, or with the good of any moral patient, for that matter. And, just as I can prudentially aim at my own good *on the whole*, I can aim at the general good *on the whole*.

This aspect of consequentialism has been subjected to vigorous attack over the years. Perhaps the best-known and most influential such attack was Rawls's (1971) criticism, which spawned the charge that consequentialism somehow neglects the "separateness of persons" (see Brink, Chapter 20, this volume):

> The most natural way, then, of arriving at utilitarianism … is to adopt for society as a whole the principle of rational choice for one man … On this conception of society separate individuals are thought of as so many different lines along which rights and duties are to be assigned and scarce means of satisfaction allocated … so as to give the greatest fulfillment of wants. The nature of the decision … is not, therefore, materially different from that of an entrepreneur deciding how to maximize his profit … or that of a consumer deciding how to maximize his satisfaction by the purchase of this or that collection of goods. … This view of social co-operation is the consequence of extending to society the principle of choice for one man, and then, to make this extension work, conflating all persons into one through the imaginative acts of the impartial sympathetic spectator. Utilitarianism does not take seriously the distinction between persons. (26–27)

Similar criticisms were urged by Gauthier (1962), Nagel (1970), and Nozick (1974). All these philosophers (and many others) are objecting to the consequentialist's embrace of interpersonal aggregation. For a consequentialist, the misfortunes of some can be outweighed by the fortunes of others. What gets a nonconsequentialist more agitated than a Promise Keeper in a Women's Studies class is the claim that the great misfortunes of a few could be outweighed by the fairly trivial fortunes of many. To use the standard example, aggregation seems to commit the utilitarian to the claim that the death of one innocent person could be outweighed by the relief of a sufficiently large number of minor headaches, and thus also to the claim that it could be permissible to kill an innocent person in order to relieve that number of headaches. Thus we must deny interpersonal aggregation.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The problem with the denial of tradeoffs or aggregation is that even committed anti-consequentialists accept them in many circumstances. For example, suppose that Homer (p. 361) is faced with the painful choice between saving Barney from a burning building, or saving both Moe and Apu from the building. Unless we want to follow John Taurek (1977), and probably Philippa Foot (1985), and possibly Judith Thomson (1997) to the funny farm,[6] we must admit that it is clearly better for Homer to save the larger number, precisely because it is a larger number. The nonconsequentialist might try to accommodate this intuition by limiting the scope of tradeoffs. For example, perhaps we are allowed to trade lives for lives (or similarly serious harms), but we are not allowed to trade lives for convenience. Homer can save the lives of Moe and Apu rather than Barney, but he can't leave Barney to die in order to provide all the inhabitants of Springfield with a few minutes extra free time every day. Tim Scanlon (1998) tries such a move in his attempt to accommodate limited aggregation in his contractualist theory:

> [I]t seems that our intuitive moral thinking is best understood in terms of a relation of "relevance" between harms. If one harm, though not as serious as another, is nonetheless serious enough to be morally "relevant" to it, then it is appropriate, in deciding whether to prevent more serious harms at the cost of not being able to prevent a greater number of less serious ones, to take into account the number of harms involved on each side. But if one harm is not only less serious than, but not even "relevant to," some greater one, then we do not need to take the number of people who would suffer these two harms into account in deciding which to prevent, but should always prevent the more serious harm.[7]

Scanlon rightly sees that it would be highly implausible to limit tradeoffs to harms of exactly equal seriousness. It is clearly better that one person suffer some particular harm than that ten people suffer a harm that is only slightly less serious. However, his attempt, or any similar attempt, to limit the scope for tradeoffs faces some serious problems. First, it is fairly clear that the relation of moral relevance does not obey transitivity. To see why, suppose, first, that it does. Consider now a descending scale of finitely many different harms, from the most serious, such as death, all the way down to the most trivial, such as a minor temporary headache. The difference in seriousness between any two adjacent harms is no larger than is necessary for the lesser harm to be clearly less serious (p. 362) than the greater harm. Suppose, also, that for every harm on the scale above the most trivial, there is some lesser harm that is relevant to it. Call this second assumption the "continuity assumption." Transitivity and continuity together entail that the most trivial harm is relevant to the most serious harm, precisely the result that the notion of moral relevance is intended to avoid. Can we preserve transitivity by rejecting continuity? This would involve finding a break (or breaks) in the scale between two harms, such that the harm directly below the break is not morally relevant to the harm directly above the break. Given that the difference between any two adjacent harms is as small as is compatible with the harms being morally distinct, the postulation of a break in the scale would run directly counter to the intuition that suggested the notion of moral relevance in the first place. Where could such a break plausibly occur? The most likely candidate would be just below death. There is, we might think, something special about death. As Clint

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Eastwood's character says in *Unforgiven*, "It takes away all a man has, and all he's gonna have." Unpleasant as even severe mutilation is, perhaps it's still worse that one person dies than that any number are mutilated. This might be the view of death espoused by those students in introductory classes who claim that life is "invaluable" or "infinitely valuable," but is it really plausible? Can anyone who really considers the matter seriously honestly claim to believe that it is worse that one person dies than that the entire sentient population of the universe be severely mutilated? Clearly not. Perhaps the break in the sequence of harms could occur at some later point. Perhaps there is some harm short of death that is worse than any number of any lesser harms. This seems even more implausible, though, than the claim that death is worse than any number of any lesser harms.

We must, therefore, conclude that the relation of moral relevance, if it is to do the work intended for it by Scanlon, does not obey strict transitivity. So what? If the notion of moral relevance were supposed to constrain our judgments of all-things-considered betterness, this would be a serious problem. Although some brave souls have seriously entertained the possibility that "all-things-considered better than" is not a transitive relation,[8] the sheer implausibility of the suggestion makes the standard objections to utilitarianism, Kantianism, or contractualism appear trivial by contrast. However, Scanlon suggests the notion of moral relevance as part of an account of what principles are reasonably rejectable, and thus of which options are permissible, obligatory, or forbidden. To demonstrate that, even in this context, the failure of transitivity leads to highly implausible results, I need to consider an example. Suppose, for the sake of argument: (a) that the loss of both arms is less serious than but morally relevant to death; (b) that a broken leg is less serious than but morally relevant to the loss of both arms, but not morally relevant to death; (c) that in a choice between saving one life and preventing one thousand **(p. 363)** people from losing both arms, it is obligatory to aid the larger group; and (d) that in a choice between preventing one thousand people from losing both arms and preventing one million people from breaking a leg, it is obligatory to aid the larger group. (The choice of examples is unimportant.) Consider now three different choices: (i) Save one life or prevent one thousand people from losing both arms. (ii) Prevent one thousand people from losing both arms or prevent one million people from breaking a leg. (iii) Save one life or prevent one million people from breaking a leg. From (b), (c), and (d) it follows that it is obligatory to aid the larger group in (i) and (ii), and the smaller group in (iii). So far, so good. But what happens when we are faced with all three options in one choice? No answer here seems satisfactory. Consider the possibility that one of the options, say saving the life, is obligatory. But now suppose that, just as you are about to save the life, it becomes impossible for you to prevent the million people from breaking a leg. Perhaps the largest group is further away than the other two, and your fuel tank is punctured by a jagged rock on the road to the one person. You are still able to save either the one or the thousand, but you can't reach the million in time. Now you find yourself faced with choice (i), in which it is obligatory to save the thousand and forbidden to save the one. But this is very strange. You were about to do your duty, virtuously eschewing both forbidden alternatives, when one of the forbidden alternatives by chance becomes unavailable, as a re-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

sult of which the other forbidden alternative becomes obligatory, and the previously obligatory alternative becomes forbidden. We should, if at all possible, avoid having to swallow such an unpalatable consequence. The same reasoning applies, mutatis mutandis, to the hypothesis that either of the other alternatives is obligatory in the three-option choice. Perhaps, then, each option is permissible in the three-option choice. But the implausibility of this can be demonstrated by the very same thought experiment. You are about to perform the perfectly permissible act of saving a life, when one of your other permissible alternatives becomes unavailable by chance. Now it is no longer permissible to save the life. A further possibility is that each option is forbidden in the three-option case. But this is even more unpalatable than the previous suggestions. Not only would we have to accept that a previously forbidden alternative can become obligatory by the chance deletion of another forbidden alternative, but we would also have to accept the existence of situations in which an agent, through no fault of her own, cannot help but do wrong. What is more, such situations may be very common. Both through the agency of charities and through our own efforts, many of us are able to bring many different types and levels of aid to others.

Up until this point, I have been ignoring a potentially important ambiguity in describing the anti-consequentialist rejection of interpersonal aggregation. The consequentialist seems to be committed both to axiological and deontic aggregation. Axiological aggregation involves the claim that harms and benefits can be traded off against each other in determining the overall goodness (or badness) of a state of affairs. Deontic aggregation involves the claim that harms and benefits can be traded off against each other in determining the relative strength of the reasons to make certain choices rather than others. Given the structure of consequentialist theories, a commitment to axiological aggregation entails a commitment to deontic aggregation. If a state of affairs with a (p. 364) large number of small benefits and a small number of large harms is better, ceteris paribus, than one with a small number of large benefits and a large number of small harms, there will always be more moral reason to choose the former over the latter. However, other ethical approaches may, at least in theory, separate the question of axiological aggregation from that of deontic aggregation.[9] If we accept at least the limited axiological aggregation that even nonconsequentialists acknowledge (for example, preferring fewer deaths to more deaths, or preferring small amounts of some harm to much greater amounts of a slightly less serious harm), is there any other plausible way to block unrestricted axiological aggregation? At first glance, it appears the answer is no. Recall the continuity assumption. It seems highly plausible that there are misfortunes that are worse than mild headaches, that nonetheless can be individually outweighed by a sufficient number of mild headaches. This is relatively uncontroversial. A mild ankle sprain is a good candidate for such a misfortune. Likewise, it’s pretty clear that there are misfortunes that are worse than mild ankle sprains, that nonetheless can be individually outweighed by a sufficient number of mild ankle sprains. Perhaps a broken ankle is such a misfortune. Even though it’s worse that one person breaks her ankle than that she mildly sprains it, it’s worse that many people have mild ankle sprains than that one has a broken ankle. But this process of escalation can be continued. For each misfortune short of the worst possible one, there

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

is a worse misfortune that can be individually outweighed by a sufficient number of the lesser one. In particular, it seems plausible that there is some misfortune short of death, perhaps some kind of mutilation, that can, if suffered by enough people, outweigh one death. Consider now a sequence of judgments, **S**, that begins as follows: one death is better than n^1 mutilations; n^1 mutilations are better than n^2 xs (where x is some misfortune less bad than mutilation). **S** continues with the first term of each comparison being identical to the second term of the previous comparison, until we reach the last two comparisons: n^{m-2} broken ankles are better than n^{m-1} mild ankle sprains; n^{m-1} mild ankle sprains are better than n^m mild headaches. If we have **S**, we can conclude, by the transitivity of "better than" that one death is better than n^m mild headaches. Thus unrestricted aggregation seems to be the only alternative to denying the transitivity of "all things considered better than."

But perhaps I have been too hasty. In reply to a paper in which I appealed to this reasoning to argue that consequentialists are indeed committed to the claim that some number of mild headaches is worse than one death, Erik Carlson (2000) produced an ingenious argument to the contrary. He suggests that something like the principle of diminishing marginal utilities might apply to harms themselves, and furthermore that there might be an upper bound to the cumulative disvalue produced by aggregating any particular kind of harm. Perhaps each type of harm, when aggregated, would asymptote toward its upper bound. Thus, for example, ten mild headaches might not be ten times as bad as one mild headache, and there may be no number of mild headaches whose aggregate disvalue is one hundred times worse than one. The exact details of the suggestion are not important, but so long as the upper bound on the disvalue of headaches falls (p. 365) short of the disvalue of one death, the continuity assumption won't license the postulation of a true **S**, the sequence that takes us all the way from a life down to headaches. To see why, consider a simplified sequence that satisfies the continuity assumption and the diminishing utility suggestion. Suppose that there are just three types of harm: headache, mutilation, and death, having disvalues of 1, 10, and 100, respectively. Suppose further that each type of harm has an upper bound on its aggregate disvalue of 15 times its individual disvalue. So no number of headaches will have an aggregate disvalue of more than 15, no number of mutilations of more than 150, and no number of deaths of more than 1,500. It is true that some number of mutilations is worse than one death, that some number of headaches is worse than one mutilation, but that no number of headaches is worse than one death.

As I said, this argument is ingenious. It is also unsuccessful. The first thing to note is that the suggestion can't simply be an application of the commonly accepted principle of diminishing marginal utilities. It may be, as a causal matter, that further headaches, or further duration of headache, becomes less unpleasant for the sufferer. My argument, though, concerns many minor unpleasant experiences, which are all (at least roughly) equally unpleasant. Furthermore, the anti-consequentialist intuition is that one death is worse than any number of minor headaches, even (or especially?) when spread out among any number of different people. It is highly implausible to suggest that the headache of the trillionth person is somehow less *unpleasant* than the otherwise identical headache of the first person, just because a whole bunch of other people have already ex-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

perienced one (or are currently experiencing one). So the suggestion must be that experiences that are equally unpleasant diminish in *badness* as they are aggregated. That is, the disutility of x headaches is less than x times the disutility of one headache.

Further, and this is a point raised by Carlson in discussing his suggestion, do we assume that the badness of harms diminishes cumulatively, starting with the first such harm ever experienced, or does the diminishing start fresh with each choice? Either option is unacceptable. If the former, your current headache may be only infinitesimally bad (if there's an upper bound to the total possible headache badness, as there would have to be to counter my argument), because countless people before you have experienced headaches. What is even more absurd, with a fine-grained enough categorization of harms, one fairly trivial harm could be worse than an intuitively much more serious one. Let me explain. Suppose that a particular mildly unpleasant nasal itch has never before been experienced. Perhaps it is only caused by a rare combination of English and French cuisine that no chef has yet been brave, or foolish, enough to attempt. Suppose the first such itch has a disutility of 2, with an upper bound of 200. Now suppose that the only way to prevent someone, say Mary, losing a leg involves producing the first instance of the nasal itch. Let's say that the loss of a leg has an initial disutility of 300, with an upper bound of 30,000. Intuitively, it is much worse to lose a leg than to experience the mildly unpleasant nasal itch. But if we apply the cumulative version of the diminishing utility suggestion, we get a strange result. Suppose that, in the course of history, so many people have lost legs that the cumulative disutility has passed 29,999. Now, each additional loss has a tiny disutility, well below 1. It now appears that it is worse for Mary to (p. 366) suffer the nasal itch than to lose a leg. But this is clearly absurd. We must, then, consider the version of the diminishing utility suggestion in which the diminishing starts fresh with each choice.

But this version also leads to unacceptable results. Consider again the simplified spectrum of harms encompassing just headaches, mutilations, and deaths. Recall that the initial disutility of a mutilation is 10, with an upper bound of 150, and the initial disutility of a death is 100, with an upper bound of 1,500. Suppose that the aggregate disutility of 2 deaths is 199, and the aggregate disutility of 2,000 mutilations is 149. According to the current interpretation of Carlson's suggestion, it is clearly worse that 2 people die than that 2,000 people are mutilated, and thus one should choose the 2,000 mutilations over the 2 deaths, if faced with the choice. But suppose also that the aggregate disutility of 1,000 mutilations is 145. In which case it is clearly worse that 1,000 people are mutilated than that one person dies, and thus one should choose the one death over the 1,000 mutilations, if faced with the choice. But now we are faced with the ridiculous possibility that we could reverse moral judgments by splitting one choice into two. Whether it's better to kill 2 than to mutilate 2,000 could depend on whether one could first choose between killing one and mutilating 1,000, and then choose between killing the other and mutilating the remaining 1,000. Let's add a few details to the example. Suppose that 2,000 people are in danger of suffering mutilation from a disease. However, if 2 other people, currently trapped in a mineshaft, die, a cure can be synthesized from their bodies that will prevent the 2,000 mutilations. You can save the two in the mineshaft by pressing a button

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

in front of you, or you can let them die. You cannot save only one. If you let them die, the 2,000 mutilations will be prevented. What is the better course of action? On Carlson's suggestion it would seem to depend on the details of how the cure will be synthesized from the two bodies. If each body can provide a cure for 1,000 people, the better course is to choose the two deaths. In effect, you are twice choosing between 1 death and 1,000 mutilations. However, if each body provides half the cure for all 2,000 people (and half a cure without the other half does no good), the better course is to choose the 2,000 mutilations. In this case, you are choosing one time between 2 deaths and 2,000 mutilations. This result is, as I said, absurd. If you set out to bring about the best state of affairs in all your choices, your decisions could differ, depending on whether you were able to split your choices up, or perhaps simply to think of your choices as split up.

Before we leave the topic of axiological aggregation, it is worth remembering that we commonly accept tradeoffs between lives and much lesser values, such as convenience. For example, we allow public projects such as building a bridge in order to make travel between two places more convenient, even when we know that several people will die in the course of the construction. Likewise, even most anti-consequentialists don't demand that highway speed limits be lowered to the optimal point for saving lives, even though the advantages of higher speed limits are increased convenience for many.[10]

(p. 367) 3. The Challenge of Value Incommensurability

Some comparisons are easy. Who has more legs, me or my cat? He does, two more, which is twice as many. Who has more hands? I do, two more, which is infinitely many more. Some comparisons are difficult, but (theoretically, at least) possible. Which lawn has more blades of grass, mine or my neighbor's? I don't know, but it would be possible, albeit tedious and time-consuming, to discover (assuming we agreed on the location of the property line). Some comparisons don't even make sense. Am I heavier than my wife is tall, and if so, by how much? We could, of course, *force* meaningfulness on that comparison, by, for example, counting (my) pounds as equal to (her) inches. But why pounds, and not kilograms, or grams, or milligrams? And why inches, and not centimeters, or millimeters, or some completely different unit of height altogether? Of course, no one is really interested in comparing my number of legs or hands with my cat's, or my lawn's count of blades of grass with my neighbor's lawn's, and certainly not my weight with my wife's height. But we are interested in lots of other comparisons, many of them involving what we think of as valuable.

For example, suppose I have a choice between two different jobs in two different locations. One of the considerations I may be interested in is my financial prospects. Suppose that job 1 is in Arselick, Indiana, and pays $50,000, whereas job 2 is in Avotoast, California, and pays $100,000. I would earn $50,000 more in Avotoast than in Arselick, but that doesn't settle the issue. No one, at least no one rational, thinks that money is *intrinsically* valuable. We value money, when we do, for what it can do for us, its *instrumental* value. It

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

may well cost me much more to live in Avotoast than in Arselick, so my overall financial prospects may be better in Arselick. Furthermore, even taking into consideration all the differences to my life that the different salaries would make in the different locations, I may value other things about the two options. Perhaps Arselick has beautiful sunsets, caused by the pollution from the nearby industrial plants and the long-running tire fire, whereas Avotoast has differently beautiful sunsets, looking out over the Pacific Ocean. Perhaps Avotoast has a large variety of excellent, but expensive, restaurants, whereas Arselick has only a few, but decent quality and affordable, restaurants. Perhaps Arselick is within easy driving distance of many of my close friends, so living there would enable me to maintain those friendships. I know no one in or near Avotoast, but the prospects for forming many new friendships (but less meaningful ones, given the reputation of Avotoasters for shallowness) are excellent there.

I have only begun to scratch the surface of the complications involved in my career choice between Arselick and Avotoast, but it is already clear that the choice between the two options could be a very difficult one to make. It is not merely tedious and time-consuming, like counting thousands of blades of grass. Because it involves what appear to be different values, or different dimensions of value, it may be very difficult, or even impossible, for me to know whether one option is overall better than the other, or (p. 368) whether they are equally preferable (or it may not—perhaps I so overwhelmingly value being able to look at the Pacific Ocean that the choice is obvious). This situation is not that rare. Life is complicated, and we aren't very good at figuring things out.[11] This can certainly pose significant practical difficulties in making either prudential or moral choices. But there is a further, and deeper, worry in the vicinity. Perhaps the differences between the intrinsically valuable aspects of some (maybe many) choices are such that it's not merely difficult, or even impossible, for us to *know* whether one choice is better than the other, and if so which one, or they are of equal value. Perhaps the options are truly incomparable, in the sense that neither is better than the other, *and* they aren't of equal, or even roughly equal, value.

Consider the following passage from Philippa Foot (1983):

> What we must ask, therefore, is whether in cases of irresolvable moral conflict we have to back both the judgement in favor of *a* and the judgement in favor of *b,* although doing *b* involves not doing *a.* Is it not possible that we should rather declare that the two are incommensurable, so that we have nothing to say about the overall merits of *a* and *b*, whether because there is nothing that *we* can say or because there is no truth of the matter and therefore nothing to *be* said … incommensurability is not an unfamiliar idea. I think, for instance, of the impossibility of saying in many cases whether one man is happier than another when one lives a quiet and contented life and the other a life that is full of joy and pain.

Foot is here suggesting incommensurability, by which she means incomparability, as an approach to what appear to be moral dilemmas, "where the application of one principle would give the judgement 'there is stronger reason morally speaking to do *a* than to do *b*'

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

and the other 'to do *b* than to do *a*.'" Let us apply this suggestion to an example, Agamemnon's tragic choice at Aulis. Considerations of utility, let us say, and Agamemnon's kingly duties rule in favor of sacrificing Iphigeneia, whereas considerations of Agamemnon's paternal duties and affections rule in favor of sparing her. The situation appears to be dilemmatic, so the suggestion goes, because the choices are in fact incomparable. We have nothing to say about the overall merits of the two choices.

Foot offers as alternative explanations of the claim that *a* and *b* are incomparable (i) that *we* can say nothing about the overall merits of *a* and *b*, and (ii) that there is no truth of the matter and therefore nothing to *be* said. Her suggestion is that cases of moral conflict involve values which we cannot compare, either because they cannot be compared (even by God), or because we are simply incapable of comparing them.

There is a big difference between (i) and (ii). A refusal to say anything about the overall merits of *a* and *b* purely on the grounds of (i) is practically troubling, but perfectly (p. 369) compatible with the view that there is something to be said (although not by us), whereas such a refusal on the grounds of (ii) is not. My concern is with this second, strong, version of incomparability, and the challenges to ethical theory that it poses.

It might be objected that Foot's suggested characterization of incomparability is misleading. The declaration, it might be argued, that *a* and *b* are incomparable *is* saying something about the overall merits of *a* and *b*. It may not be saying anything which is of much use in trying to decide what to do, but it is still saying something. Perhaps we should recast Foot's alternatives as follows: (i') We have nothing to say about whether *a* is better than *b*, *b* is better than *a*, or they are of equal value; (ii') It is not the case that *a* is better than *b*, it is not the case that *b* is better than *a*, and it is not the case that they are of equal value.

Perhaps (ii') is really what is meant by (ii). But the application of either (ii) or (ii') to a situation involving a choice between *a* and *b* gives the result that morality cannot, even in principle, guide our actions in this case. Whichever option we choose, morality will have nothing to say about whether we made the better or worse choice.

In a similar vein to Foot's suggestion, Joseph Raz (1986) also appeals to the notion of incomparable values to explain the existence of moral dilemmas, in which an agent can't help but act wrongly.

> Incommensurability[12] may help in the explanation of [moral dilemmas] … If all the agent's options involve wronging others or just doing evil and if none of them is the lesser evil then there is no action that is the right action for him to do. … It is true that this is not a case where the agent will fail in performing the right action, not a case in which his act will be worse than some other option open to him. … It is question-begging to assume that a person can only do wrong by failing to perform a better act which he could have done. … Incommensurability shows that

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> there is conceptual room for a notion of wrongdoing which does not involve failing to take a better action available to one. (359–360)

Raz defines incommensurability as follows:

> A and B are incommensurate if it is neither true that one is better than the other nor true that they are of equal value. (322)

Raz gives, as what he calls "the mark of incommensurability," the failure of transitivity, which distinguishes incommensurability from equality.

> Two valuable options are incommensurable if (1) neither is better than the other, and (2) there is (or could be) another option which is better than one but is not better than the other. (325)

Raz appears to be talking about incomparability in the sense of my interpretation (ii) or (ii') of Foot's suggestion.

(p. 370) We are now in a position to examine what I will call the "strong" sense of incomparability. Let us say that options *a* and *b* are strongly incomparable when it is not the case that *a* is to be preferred over *b*, *b* over *a*, or they are to be accorded equal weight. The claim here is not merely an epistemic one. In cases of strong incomparability, even God could not judge that one option was morally preferable to the other or that they were of equal moral value. In such a situation there would be no correct moral comparison between the two options, and thus morality could be of no use in deciding between them.

Why would anyone accept strong incomparability (henceforth simply incomparability)? Anyone who is disposed to accord basic moral value to such diverse features of actions and persons as pleasure, truth, courage, integrity and self-fulfillment, might well regard such goods as incomparable, because they appear to be such different features of persons, actions, or states of affairs. We experience the difficulty, perhaps verging on impossibility, of judging whether a particular display of courage is morally better than a particular display of honesty, and consider as the best explanation for this difficulty the claim that neither is better than the other nor are they equal in value. Of course, it might appear that hedonists about intrinsic value can avoid this conclusion. If truth, courage, integrity, and so on don't really have *intrinsic* value, but are only valuable to the extent that they stand in some relation (causing? promoting?) to pleasure, happiness, or the like, we might think that the difficulties of making the relevant comparisons just amount to the difficulty of knowing, for example, how much pleasure the different displays of courage and honesty will promote. But this would be a mistake. Hedonism may well be the correct theory of intrinsic value, but any plausible version of hedonism will have to allow that both pleasures (broadly construed) and pains (equally broadly construed) are radically heterogeneous. Foot's earlier example to illustrate her hypothesis seems to be one of comparing different lives hedonically: "I think, for instance, of the impossibility of saying in many cases whether one man is happier than another when one lives a quiet and contented life and the other a life that is full of joy and pain." Try to compare the pleasure of

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

watching an excellent movie with the pleasure of eating a delicious meal. It's easy to say that both experiences are extremely pleasurable, but we often don't know how to begin to judge that one is more pleasurable than the other, or that they are equally pleasurable. Perhaps the best explanation for the difficulty is that the two pleasures are incomparable. Embracing hedonism, it seems, won't magically eliminate the threat of incomparability.

It might be thought that incomparability will only be a problem in a relatively small number of cases, and that in most cases we will be able to make a comparison between two different states of affairs, say, which contain heterogeneous goods. To return, for the sake of illustration, to a nonhedonistic value theory, we might value both truth and pleasure, recognize that there will be problem cases where only a lie can bring about a certain amount of pleasure, and hold that in some of these cases the options of lying to bring about the pleasure and telling the truth and failing to bring it about are incomparable. We might also think that there will be cases where it is obviously better to tell a small lie to bring about a large amount of pleasure (or avoid a large amount of pain), and, conversely, where it is obviously better to forgo a small amount of pleasure (or suffer a (p. 371) small amount of pain) than to tell a heinous lie. To use a familiar (though misleading) schematic example, we might say that nineteen apples are obviously better than three oranges, but that four apples and three oranges are incomparable.

Call the view that incomparability only applies in a small number of comparisons "Limited Incomparability." At first sight, this might seem plausible. Recall my earlier example of the comparison between viewing and dining experiences. If I try to compare the experience of watching a fine movie with the experience of eating a small dish of peanuts, I have no difficulty in judging that the former is better than the latter. Likewise, if I compare the experience of watching a moderately amusing twenty-two-minute sitcom (take your pick), with the experience of eating an excellent meal, I have no difficulty in judging the latter to be better than the former. But, if we look more closely at Limited Incomparability, complications arise.

Return to the schematic apples and oranges comparisons. The suggestion is that four apples and three oranges are incomparable, but that nineteen apples are better than three oranges. But now, what happens if, starting with the original comparison of four apples with three oranges, we gradually increase the number of apples, while holding the number of oranges steady? For each different number of apples either there will or there won't be a fact as to whether that number of apples is better than three oranges. Presumably there will be a smallest number of apples such that that number of apples is better than three oranges. Let us suppose that number to be seven. Seven apples are better than three oranges, but six apples and three oranges are incomparable. If we now increase the number of oranges to, say, five and vary the number of apples, we should be able to discover the smallest number of apples such that that number of apples is better than five oranges. Let us suppose that number to be eleven. Eleven apples are better than five oranges, but ten apples and five oranges are incomparable. This gives us an interesting result. We can move from one pair of numbers of apples and oranges, such that that

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

number of apples is the smallest number of apples which is better than that number of oranges, to another such pair by adding to the first pair four apples and two oranges.

So what is the significance of this result? It seems reasonable to suppose that something like the following principle is true for any two measures x and y:

> SD: If the difference between n^1x and m^1y is the same as that between n^2x and m^2y ($n^1x - m^1y = n^2x - m^2y$), then $n^2x - n^1x = m^2y - m^1y$. (Where $n^2 > n^1$ and $m^2 > m^1$)

Let us define a relation between amounts, a^1 and a^2, of values, x and y, such that a^1x is *minimally better* than a^2y as follows:

> MB: a^1x is minimally better than a^2y if a^1x is better than a^2y and any amount of x less than a^1x would not be better than a^2y.

This assumes that the values in question are discreet (though perhaps extremely fine-grained), rather than continuous. This seems plausible, especially if some version of hedonism is true (there will be smallest perceptible differences), but the postulation of (p. 372) continuous values doesn't present a problem. It is possible to adapt MB to incorporate the notion of *minimal significant betterness*. I leave that, and the relevant adaptations of other principles and results, as exercises for the reader.

Let us now apply these principles to the apples and oranges example. What we discovered was that seven apples are minimally better than three oranges and that eleven apples are minimally better than five oranges. The difference, that is, between seven apples and three oranges is the same as that between eleven apples and five oranges. If we apply SD to this finding, we get the result that four apples and two oranges are of equal value.

It is possible, then, to take any number of apples (or oranges) and to apply this technique to discover what number of oranges (or apples) is equal in value to it. But if, for any number of apples, there is a number of oranges which is equal in value to it, for any number of apples, there is *no* number of oranges which is incomparable in value with it. Suppose it is suggested that n apples and m oranges are incomparable. There is a number of oranges, m', such that n apples and m' oranges are equal in value. m oranges are either better, worse, or equal in value with m' oranges. But then m oranges are either better, worse, or equal in value with a number of oranges which is equal in value with n apples. There is no room for incomparability.

A defender of limited incomparability might object that although seven apples are minimally better than three oranges, and eleven apples are minimally better than five oranges, this does not give the result that the difference between seven apples and three oranges is the same as the difference between eleven apples and five oranges, at least not in the sense of "same difference" used in the principle SD. There is another sense of "same difference," such that the difference between n^1 and m^1 may be the same as that between n^2 and m^2 without it being the case that $n^1 - m^1 = n^2 - m^2$. It may, for example, be said that the difference between five and ten is the same as that between ten and twenty.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Ten is twice five and twenty is twice ten. Thus, the objection may continue, the difference between seven apples and three oranges may only be the same as the difference between eleven apples and five oranges in the sense that seven apples are better than three oranges in the same proportion as eleven apples are better than five oranges.

This objection fails to defend limited incomparability, though. If seven apples are better than three oranges in the same proportion as eleven apples are better than five oranges, there must be a specific proportion such that both seven apples are better than three oranges by that proportion and eleven apples are better than five oranges by that proportion. This will give us thoroughgoing comparability. Let us suppose that, for two values A and C, instance A^1 of A is claimed to be incomparable with instance C^1 of C, but that it is also admitted that instance C^2 of C is minimally better than A^1. If the relation of minimal betterness is proportional, there is a proportion m:n such that C^2 is better than A^1 in the proportion m:n. There is also a possible instance of C, C^3, such that C^2 is better than C^3 in the proportion m:n. It follows that C^3 is equal in value with A^1. (If C is a discrete value, C^2 may be better than C^3 only approximately in the proportion m:n. In which case, C^3 will be only approximately equal with A^1. This does not affect the argument.) Since C^3 is an instance of the same value as C^1, it is either better, worse, or equal in value (p. 373) to C^1. But then, since A^1 is equal in value with an instance of C which is either better, worse, or equal in value with C^1, A^1 must also be either better, worse, or equal in value with C^1. Once again, there is no room for incomparability.

It might be objected that my example of apples and oranges relies on the fact that amounts of apples and oranges are naturally expressed in numbers. I may not find it so easy to get the result I want if I talk about values that don't lend themselves so readily to numerical quantification. Let's consider fatigue and boredom. What if someone were to claim that some episodes of fatigue and boredom are comparable, but that others are not? Yesterday afternoon's episode of fatigue, for example, was clearly worse than yesterday morning's episode of boredom. This afternoon's episode of fatigue, on the other hand, is incomparable with this morning's episode of boredom. All episodes of boredom are comparable with all other episodes of boredom, and all episodes of fatigue are comparable with all other episodes of fatigue. Given an episode of fatigue, F^1, and an episode of fatigue, F^2, such that F^2 is worse than F^1, there is an episode of fatigue F^3 such that it is equally bad to experience both F^1 and F^3 as to experience F^2. The same holds for episodes of boredom. We may now state SD as it applies to boredom and fatigue:

> SD_{bf}: If an episode of fatigue, F^1, is better than an episode of boredom, B^1, by a certain amount, and an episode of fatigue, F^2 is better than an episode of boredom, B^2, by the same amount, (and F^2 is worse than F^1), then any episode of fatigue, F^3, such that experiencing F^3 in addition to experiencing F^1 is equally as bad as experiencing F^2 is equal in value with any episode of boredom, B^3, such that experiencing B^3 in addition to experiencing B^1 is equally as bad as experiencing B^2.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

We can also apply MB to episodes of boredom and fatigue as follows: an episode of fatigue, F^1, is minimally better than an episode of boredom, B^1, if and only if experiencing F^1 is better than experiencing B^1, and experiencing any episode of fatigue worse than F^1 would not be better than experiencing B^1.

If it is suggested that some episodes of boredom are comparable with some episodes of fatigue, but that some episodes of boredom are also incomparable with some episodes of fatigue, in particular that F^1 and B^1 are incomparable, we can use SD and MB to establish that F^1 is equal with some episode of boredom. Consider an episode of boredom, B^2, that is minimally better than F^1. There is an episode of boredom, B^3, such that experiencing B^3 and experiencing B^2 is minimally better than experiencing F^1 and experiencing another episode of fatigue, F^2, that is equally bad to experience as F^1. By SD, B^3 is equal to F^1 (and to F^2). Since episodes of boredom are comparable with other episodes of boredom, B^3 is either better, worse, or equal in value with B^1. So F^1 is equal in value with an episode of boredom that is either better, worse, or equal in value with B^1. So B^1 and F^1 are not incomparable.

Perhaps it will be objected that my argument fails for boredom and fatigue, because it is not the case that all episodes of boredom are comparable with all other episodes of boredom. Episodes of boredom are too heterogeneous to be thoroughly comparable. Some episodes of boredom are incomparable with other episodes of boredom. Perhaps boredom (p. 374) is a genus with many different species. But if this is so, we can reapply my argument to the species, and establish comparability for the genus. What if the species do not display thoroughgoing comparability? We move to the next level down and try again. What if there is no level which displays thoroughgoing comparability? I find this hard to imagine. It would require that there be episodes of a particular species of boredom such that there are no possible episodes of the same species equal in value. Equality is the building block of comparability.

Perhaps, though, there is a further problem with my simplified schematic example, involving the highly coarse-grained schematic values of apples and oranges. When I considered increasing the number of apples to compare with three oranges, I said, “For each different number of apples either there will or there won’t be a fact as to whether that number of apples are better than three oranges.” This might seem plausible for apples and oranges, but what about actual values that are more fine-grained (or maybe even continuous)? Consider again the claims that this morning’s episode of boredom, hence simply Boredom, and this afternoon’s episode of fatigue are incomparable, but that some episodes of fatigue are better than Boredom. In applying MB and SD_{BF} to this case, I assumed that for each decreasingly bad episode of fatigue (starting with this afternoon’s episode), it was either true or false that the episode was better than Boredom, so that we could, in principle at least, discover the episode of fatigue that was minimally better than Boredom. But what about indeterminacy (or vagueness—for my purposes here, the two can be treated alike)? Perhaps there are some episodes of fatigue such that it is indeterminate whether they are better than Boredom. Fine. But then we could postulate the relation of *minimal determinate betterness*. An episode of fatigue is minimally determinately

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

better than Boredom, just in case it is determinately better than Boredom, and any episode worse than it would not be determinately better. But now the defender of indeterminacy could simply move up a level and claim that, for some episodes of fatigue, it may be indeterminate whether they are determinately better than Boredom. The obvious answer to this is to postulate the relation of *determinate determinate betterness*. But then we move up another level, and so on. It might appear that, for any level or kind of determinacy, it might be indeterminate whether an episode of fatigue is better than Boredom to that level or kind. But this would be incorrect. Consider the relation of *superdeterminate betterness*. An episode of fatigue is superdeterminately better than Boredom, just in case it is better than Boredom, and there is no indeterminacy at any level about whether it is better than Boredom. Could the defender of indeterminacy claim that some episodes of fatigue are such that it is indeterminate whether they are superdeterminately better than Boredom? Obviously not. The postulation of indeterminacy about superdeterminacy is self-defeating. We can, then, replace the relation of minimal betterness with the relation of minimal superdeterminate betterness, and the rest of the argument goes through. Indeterminacy lends no support to limited incomparability.

There is a further complication to deal with. My argument against Limited Incomparability makes use of the claim that, if *a* and *b* are equal in value, and *c* is better (worse) than *a*, then *c* is better (worse) than *b*. The assumption is that the only three comparison relations are "better than," "worse than," and "equal to" (though, in fact, all we need is (p. 375) two relations, plus the assumption that, if *a* is better than *b*, then *b* is worse than *a*). Some philosophers, such as Ruth Chang (2002) and Joseph Raz (1986), suggest a fourth relation, "parity" (Chang) or "rough equality" (Raz), partly defined by a failure of transitivity. As I explained earlier, Raz claims the failure of transitivity to be what he calls "the mark of incommensurability" (by which he means incomparability). For Raz, two options may be both incomparable and roughly equal. Here is his example:

> Here are two cups, one of coffee and one of tea. As it happens (a) neither is of greater value to me than the other; (b) warming the cup of tea a little will improve its value; and (c) the improved cup of tea will be neither better nor worse than the cup of coffee. (328–329)

For Chang, on the other hand, parity is consistent with comparability. One of her examples is a comparison between the creativity of Mozart and Michelangelo. We may refrain from judging that either was better than the other with respect to creativity, admit that a little bit more creative version of Michelangelo, Michelangelo^{+} would have been better than Michelangelo, but still not consider Michelangelo^{+} to be better than Mozart. Rather, we would judge them to be on a par.

What should we say about this suggestion, that there is a relation of "parity" or "rough equality," distinguished by a failure of transitivity? Consider one further example. Suppose I have a choice between a glass of Russell Syrah (2013) (R13) that has been given time to breathe in a crystal decanter and a Perfect Manhattan (PM), made with an excellent small-batch bourbon, high-quality dry and sweet vermouth, three of those fancy cher-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ries you have to order online (not cheap maraschino crap), orange bitters, and chilled as much as possible (but not watered down with ice), and served in a chilled cocktail glass. It would be eminently reasonable to be both highly enthusiastic about both drinks and indifferent between them. Now consider a slightly less perfect Perfect Manhattan, PM⁻. Perhaps it is fractionally less chilled, or the bourbon is just ever so slightly less excellent. Still, it is probably the best Perfect Manhattan that you will ever drink. It may be clear that PM is better than PM⁻, but that it is still reasonable to be indifferent between R13 and PM⁻. Of course, we can keep imagining ever less perfect Perfect Manhattans, until we get to $PM^{holycrap}$, made with Old Crow, cheap vermouth, hideously unnaturally colored cheap supermarket maraschino cherries left over from the sundae bar at a children's party, served room temperature in a plastic cup. Long before we get to $PM^{holycrap}$, in fact, well before we get to $PM^{prettydecent}$, it becomes clear that R13 is better than it. So what should we say about the claim that R13 and PM (or PM⁻) are on a par, or roughly equal, but not equal in value? In light of the arguments I have presented, it's clear that the notions of parity or rough equality are plausible as practical strategies, given our cognitive limitations (Chang's discussion suggests something like this), but that they don't correspond to a basic comparison relation in addition to better (worse) than, and equal to. It makes sense for us to be indifferent between R13 and PM, and between R13 and PM⁻, while maintaining a preference for PM over PM⁻, because it is simply more difficult for us to compare an excellent glass of Syrah with an excellent Perfect Manhattan. (p. 376) We should admit that at most one of PM and PM⁻ is really equal in value with R13, that we can't tell which (if either), and that it is reasonable for us to behave as if they are both equal in value with R13, when given a binary choice.

It seems, then, that Limited Incomparability, as a thesis about the objective nature of moral values, is untenable. If there is strong incomparability between values (or between valuable options), it is thoroughgoing and pervasive. Far too thoroughgoing and pervasive to be plausible. The claim that it isn't objectively better for my finger to be scratched than for the whole world to be destroyed might appeal to Hume or Foot,[13] but it is obviously false. As a thesis about the limitations of our abilities to make value comparisons, Limited Incomparability makes sense, but poses no threat to any ethical theory. It is fairly common for students (and simple-minded colleagues) to cite the difficulty of making such comparisons as a reason to believe that consequentialist theories, such as utilitarianism, are false. "How can we compare my happiness with yours?," they cry. Or "How can we compare the pleasure I get from eating meat with the suffering of the animals who provide it?" In some contexts, questions like these are reasonable to ask about how to apply particular ethical theories. But they have no relevance to the question of whether any of those theories are true. There is no theory (with the possible exception of the intellectually lazy, anti-theory of moral particularism) that doesn't face difficulties of application. Is there a god, and if so, what does it really command? What would the ideally virtuous person actually do? How do we weigh prima facie duties? And has anyone, including Kant himself, ever really had a clue how to apply the categorical imperative?

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

4. Conclusion

I have argued that value comparability, either in the form of the kind of quantitative comparability involved in aggregation (either intrapersonal or interpersonal), or in the form of the kind of qualitative comparability involved in comparing putatively different values, does not threaten the theoretical soundness of consequentialist theories. Unrestricted axiological aggregation (intrapersonal or interpersonal) is supported by overwhelmingly plausible assumptions about ordinary value comparisons, combined with the near-conceptual truth that "all things considered better (or worse) than" obeys transitivity. Although it might be tempting to think that qualitatively different values are simply incomparable, when focusing on difficult choices, the overwhelmingly plausible view that comparability kicks in at extremes leads to the conclusion that comparability must also obtain in the difficult cases. This is all consistent with the view that value comparability can sometimes, perhaps even often, pose difficulties of application for consequentialist theories. It may often be difficult to compare choices that require quantitative or qualitative value comparisons. But all ethical theories face difficulties of application, many far worse than those posed to consequentialism by value comparability.

References

Carlson, Erik. 2000. "Aggregating Harms—Should We Kill to Avoid Headaches?" *Theoria* 66, no. 3: 246–255.

Chang, Ruth. 2002. "The Possibility of Parity." *Ethics* 112, no. 4: 659–688.

Foot, Philippa. 1983. "Moral Realism and Moral Dilemma." *The Journal of Philosophy* 80: 379–398.

Foot, Philippa. 1985. "Utilitarianism and the Virtues." *Mind* 94: 196–209.

Gauthier, David. 1962. *Practical Reasoning: The Structure and Foundations of Prudential and Moral Arguments and their Exemplification in Discourse.* Oxford: Oxford University Press.

Hume, David. (1739). 2000. *A Treatise of Human Nature.* Oxford: Oxford University Press.

Nagel, Thomas. 1970. *The Possibility of Altruism.* Oxford: Oxford University Press.

Norcross, Alastair. 1997a. "Comparing Harms: Headaches and Human Lives." *Philosophy & Public Affairs* 26: 135–167.

Norcross, Alastair. 1997b. "Good and Bad Actions." *The Philosophical Review* 106, no. 1: 1–34.

Norcross, Alastair. 1999. "Intransitivity and the Person-Affecting Principle." *Philosophy and Phenomenological Research* LIX, no. 3: 769–776.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Norcross, Alastair. 2002. “Contractualism and Aggregation.” *Social Theory and Practice* 28, no. 2: 303–314.

Norcross, Alastair. 2006. “Reasons without Demands: Rethinking Rightness.” In Blackwell Contemporary Debates in Moral Theory, edited by James Dreier, 38–53. Oxford: Wiley-Blackwell.

Norcross, Alastair. 2009. “Two Dogmas of Deontology: Aggregation, Rights, and the Separateness of Persons.” *Social Philosophy & Policy* 26, no. 1: 76–95.

Norcross, Alastair. 2020. *Morality by Degrees: Reasons without Demands.* Oxford: Oxford University Press.

Nozick, Robert. 1974. *Anarchy, State, and Utopia.* New York: Basic Books.

Quinn, Warren. 1990. “The Puzzle of the Self-Torturer.” *Philosophical Studies* 59: 79–90.

Rachels, Stuart. 1998. “Counterexamples to the Transitivity of ‘Better Than.’” *Australasian Journal of Philosophy* 76, no. 1: 71–83.

Rawls, John. 1971. *A Theory of Justice.* Cambridge, MA: Harvard University Press.

Raz, Joseph. 1986. *The Morality of Freedom*. Oxford: Oxford University Press.

Scanlon, Thomas. 1998. *What We Owe to Each Other.* Cambridge, MA: Harvard University Press.

Sidgwick, Henry. (1907). 1981. *The Methods of Ethics* (7th ed.). Indianapolis: Hackett.

Taurek, John. 1977. “Should the Numbers Count?” *Philosophy and Public Affairs* 6: 293–316.

Temkin, Larry. 1987. “Intransitivity and the Mere Addition Paradox.” *Philosophy & Public Affairs* 16: 138–187.

Temkin, Larry. 1996. “A Continuum Argument for Intransitivity.” *Philosophy & Public Affairs* 25: 175–210.

Thomson, Judith Jarvis. 1997. “The Right and the Good.” *The Journal of Philosophy* 94, no. 6: 273–298.

Notes:

([1]) The qualifier “net” as applied to happiness, or good, indicates that both happiness and unhappiness, or both good and bad, have been taken into account. Henceforth, I will omit the qualifier, but it should be understood, unless explicitly ruled out.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([2]) This account is, as I said, rough, because of the possibility that two actions equally maximally promote net happiness. In which case, either would be right, but neither would be obligatory. They would both be permissible.

([3]) Satisficing theories reject maximization as the criterion of rightness.

([4]) Scalar theories reject the deontic notions of rightness, wrongness, permissibility, obligation, and the like as being fundamental aspects of consequentialism at all. See Norcross (1997b), Norcross (2006), and Norcross (2020).

([5]) Most consequentialists, of course, extend moral consideration to sentient nonhuman animals. For simplicity of exposition, I will focus on examples involving people. Everything I say is equally applicable to all moral patients.

([6]) Taurek argues for the radical thesis that there is no reason to prefer the death of one person over the deaths of five persons (or even of five million persons), because there is no sense in which it is a worse thing in itself (as opposed to from the perspective of one or another of the individuals involved) for five persons (or five million persons) to die than for one person to die. Foot argues that no sense can be made of one state of affairs being overall better than another from the perspective of morality. Foot's position seems to be in agreement with Taurek's in the following sense: it rejects the claim that I have a moral reason to prefer the death of one to the deaths of five, if that reason is supposed to be grounded in the claim that it is overall better that only one person dies than that five persons die. Thomson's position is less clear. She criticizes utilitarianism for its reliance on comparative judgments of the goodness of states of affairs, and in this respect seems to be sympathetic to Foot's position. However, a charitable reading of her article (which she would no doubt reject) renders it as a defense of rule utilitarianism.

([7]) For a more comprehensive critique of Scanlon's attempts to accommodate limited aggregation, see Norcross (2002).

([8]) For examples of the attempt to deny transitivity for "all-things-considered better than," see Temkin (1987; 1996); Quinn (1990); and Rachels (1998). Temkin (1996) uses the same central example as Rachels (1998) (which was written earlier), but Temkin's explanation for the supposed intransitivity is the same as the one he provides in his 1987. Quinn doesn't explicitly claim that "better than" is intransitive, but his arguments, if successful, entail that a utilitarian should deny the transitivity of "better than." I discuss Temkin (1987) in Norcross (1999). I discuss Temkin (1996) and Quinn (1990) in Norcross (1997a).

([9]) For criticisms of nonconsequentialist attempts to block deontic aggregation, see Norcross (2009).

([10]) For detailed discussion of both these points, see Norcross (1997a).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([11]) It might appear that what I'm saying here is in tension with what I argued in the previous section. In that section, I argued that it is possible to aggregate many small harms to outweigh some large harm(s), and that we often do so. My argument there concerned truths in axiology. My claim here is an epistemic one. The epistemic point might be thought by some to motivate an ontological claim. But the difficulty we have in *making* value comparisons is perfectly compatible with there being truths about such comparisons, as I will argue.

([12]) Raz uses "incommensurability" to mean "incomparability."

([13]) See Hume (2000) and Foot (1985).

Alastair Norcross

Alastair Norcross is Professor of Philosophy at the University of Colorado Boulder, where he has taught since 2007. Prior to that, he taught at Southern Methodist University and Rice University (before being allowed out of Texas for good behavior). He works both on ethical theory and on issues in applied ethics. In ethical theory he has published extensively on consequentialism, in particular defending a scalar version of the theory. His book Morality by Degrees: Reasons without Demands (Oxford University Press, 2020) articulates and defends the scalar approach. In applied ethics he has published many articles criticizing the common practices of raising animals for food and using them in experimentation, including the widely reprinted "Puppies, Pigs, and People: Eating Meat and Marginal Cases" (Philosophical Perspectives, 2004). He also runs marathons, with somewhat less success than Eliud Kipchoge, and writes, directs, and acts in the theater, with somewhat less success than Kenneth Branagh.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

What Should a Consequentialist Promote?

Katarzyna de Lazari-Radek

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.4

Abstract and Keywords

Whether we choose to be act or rule consequentialists, to maximize value or be satisfied with a less demanding requirement, and even whether we require impartiality in our values, we need to decide what set of values we should commit ourselves to. In this chapter I will ask what intrinsic good or goods a consequentialist should value. I will start with a general point: what reasons are there for being a value monist rather than a value pluralist? Then, I will ask whether consequentialists should be concerned with what is good for someone or rather with what is good "for the world." To answer this, I will discuss how best to understand the notion of welfare. Finally, I will consider different welfarist theories and suggest a possible candidate for the most promising one.

Keywords: intrinsic good, monism, pluralism, well-being, welfare, resonance, hedonism, pleasure

1. Introduction

MANY of the contributions to this volume are, in one way or another, responding to the question: "How should a consequentialist bring about good consequences?" Should we maximize consequences or merely satisfy some lesser requirement? Should we guide our actions by rules that, if followed, will produce the consequences we value, or should we perform the acts that will produce them directly? Is there a place for supererogation in consequentialism or is whatever we do situated within a realm of obligation? How impartial are we supposed to be in practice? Can we give priority to some people? Or to our fellow humans? No matter how we answer these questions, we still need to decide on the substance of the good consequences, that is, on the value or values we should bring about.

Though certain value theories are strongly associated with some forms of consequentialism (as, for example, hedonism and desire-based theories are associated with utilitarianism), it seems reasonable to assume that being a consequentialist is compatible with any values. Philip Pettit has defended this claim, suggesting that there are no particular consequentialist-like values that a consequentialist needs to limit herself to. In his opinion,

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

the difference between consequentialist and nonconsequentialist theories is not about the question of what kind of value to bring about but rather how to act in regard to that value. Pettit argues that a consequentialist is concerned with "promoting" a value rather than "honoring" it in our own lives and actions.[1] Thus, he finds nothing strange in a consequentialist "promoting a value as (p. 198) intimately linked with non-consequentialist theories as that of respect for persons."[2] "[T]o promote this," he continues, "will be to try to ensure that people respect one another as much as possible, even if this requires disrespecting some."[3] For the purpose of this paper, I am going to accept this claim.

If so, however, a chapter like this one, on values, could stand on its own and be included in a companion to any moral theory. After all, whether you are a consequentialist, a deontologist, a virtue theorist, or a proponent of an ethics of care, you need some axiology to make choices about the good and the bad. I will hope to justify the importance of this chapter in this volume by addressing problems that consequentialists must face. First, I will consider whether consequentialists should be monists or pluralists. Second, I will ask whether they should limit the values they seek to promote the welfare of sentient beings or should accept that there are values beyond any being's welfare. To answer this question, I will discuss different understandings of well-being and look for the most appealing one. Third, I will raise the normative question whether consequentialists should promote any particular value or values.

2. Intrinsic Value—A Preliminary Point

Philosophically significant discussions about the good commonly focus on the issue of intrinsic value. A consequentialist could promote all sorts of goods in the world: from social justice to the number of cats in our houses. Whatever she chooses though, she must decide whether the chosen good is a means to something more important or is good in itself. Things that are good only as a means to something else are good instrumentally. Things that are good in themselves are good intrinsically; they are worthy of our effort for their own sake, and not because of the consequences they bring about.[4] It is about this latter kind of good that philosophers are most likely to disagree.

2.1. Should a Consequentialist Be a Monist or a Pluralist?

Let us begin with the most basic question: how many intrinsic goods are there for a consequentialist to promote?

(p. 199) All the classical utilitarians, Jeremy Bentham, John Stuart Mill, and Henry Sidgwick, believed that there is only one such good—happiness, which they defined as pleasure. Thus, they were value monists.[5] Monistic consequentialists claim that there is only one intrinsic or ultimate good. In *The Methods of Ethics*, Sidgwick argued that pleasure, understood as a particular kind of consciousness, is the only intrinsic value; he stated that all other goods like truth, knowledge, beauty, or freedom are valuable only in so far as they bring about the intrinsic value of pleasure. "It is paradoxical," he wrote, "to maintain that any degree of Freedom, or any form of social order, would still be commonly re-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

garded as desirable even if we were certain that it had no tendency to promote the general happiness."[6] Pleasure is the only value that is worth pursuing for its own sake.

Since the second half of the twentieth century, another monistic theory has become popular among consequentialists: a desire theory. R. M. Hare, for example, defended a view that can be seen as implying that the right act is the one that maximally satisfies the preferences of all those affected by it.[7] Many economists have also assumed that some form of desire theory is the best way to understand welfare.[8]

Sidgwick's student, G. E. Moore, thought that Sidgwick's intuition that there is only one intrinsic good and that only states of consciousness can be intrinsically good, clashes with our other intuitions. Imagine that you have a choice, he argues, between two worlds—one full of beautiful plants, rivers, mountains, and so on, and the other "the ugliest world you can possibly conceive … simply one heap of filth."[9] Now imagine that no one will ever experience these worlds at all. Is it better that one of these worlds should exist, rather than the other? Moore's intuitions told him: yes, it is better that the first one should exist. This in turn made him believe that there are other things to cherish and search for apart from pleasure alone, and even apart from any states of consciousness. In particular, he thought that beauty and friendship are good in themselves. Hedonism cannot be right, he stated, in limiting everything of value to mental states, and to only one type of mental state—pleasure.

Thus, Moore was a value pluralist. Pluralism assumes that there are several irreducible goods. The fact that goods are irreducible means that we cannot use one of them to justify the claim that something else is an intrinsic good. Each of them stands on its own. Pluralists present varying lists of what is intrinsically good. Derek Parfit, for example, includes having knowledge, engaging in rational activity, experiencing mutual love, and (p. 200) being aware of beauty.[10] William Frankena is well known for what might be the longest list of intrinsic values written by a philosopher; just to enumerate a few from his list: life, consciousness, health, pleasures, happiness, truth, knowledge, beauty, love, freedom, peace, novelty, good reputation, and honor.[11]

The obvious advantage of pluralism is its consistency with our common-sense intuitions about value.[12] When we wonder what makes our lives successful, we usually think of a few different/common goods: a close family, knowledge, freedom, happiness, being a moral person, and so on. The hedonist's choice may seem like an indefensible limitation. Each list proposed by pluralists, however, faces the same objection. If pluralists enumerate different values that are not substantially similar, the natural question seems to be: why these values and not any other ones? On what basis are these ones chosen? In order to defend listing them all as intrinsic values, we may look for something that justifies our choice. If, however, we give a reason for our choice, doesn't this imply that this reason points to another intrinsic value? That would turn value pluralism into monism.

Against this objection, a pluralist can defend her position in at least three ways. First, she may claim that it is the distinctive feature of a pluralist theory that the listed goods are different and cannot be justified by a single criterion. She can maintain that we have a va-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

riety of values, each of which has a different property that makes it valuable in itself. Second, she can appeal to a specific method of discovering true judgments about values—say, intuition. She may argue that what is good in itself can be known only via intuition and it is impossible to provide any further explanation of why these and not other values are intrinsic. Finally, she may put a similar demand to a monist: why does a monist choose this good, and not any other good? If he presents a justification, it can become an intrinsic value, and so ad infinitum. A monist can, of course, appeal to our intuitions just as a pluralist has done but then he can scarcely ask a pluralist "why these values?"

A more significant objection to pluralism is that conflicting multiple values cannot guide actions. It seems plausible to hold that if we accept something as a value, we are acquiring a reason to act so as to promote it. If there are multiple goods, however, or values that have no common measure and cannot be reduced to any single value, then if they conflict and we cannot promote them all, how can we know which value or values to promote?

Imagine that you are a value pluralist consequentialist who believes that there are two intrinsic values: pleasure and knowledge. Tonight, you have to choose between the following options: either you will watch a romantic comedy with your family, which will bring you value in the form of pleasure or you will read a philosophical book, which will (p. 201) increase your knowledge and improve your intellectual abilities. If both pleasure and knowledge are of the same importance, which should you choose?[13] You can decide to promote one value tonight and the other one tomorrow; or, perhaps, you should choose the value that you will get the most of—for example, if you get a lot of pleasure from romantic comedies, but only a little knowledge from philosophical books, you will choose the former and not the latter. Can we really quantify pleasure and knowledge, in the manner suggested by this example? Moreover, what if you always get more of one of the values and not the other? Are you allowed to opt for one of the values all the time? Can you neglect some of the values as you bring about some others? Should you promote them equally or according to some other rule? The situation becomes even more complicated if you are a pluralist who believes not in two but, let us say, ten different values?

Bernard Williams famously stated that there is no common measure that allows us to compare values and determine gains and losses.[14] Values are "incommensurable," which means that we are forced to face value conflicts that cannot be resolved by rational considerations. Ruth Chang develops Williams's claim into her own theory of incomparability.[15] Two values are "incomparable" if "they fail to stand in an evaluative comparative relation, such as being better than or worse than or equally as good as the other."[16] Chang argues that this leads to a serious problem for a pluralist consequentialist:

> As many philosophers believe, you're justified in choosing one alternative over another only if it is better or as good as the other, and incomparability holds when it's false that they stand in any such comparative relation. Incomparability among alternatives, then, leads to a breakdown in practical reason. If incomparability is

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> widespread, then what we do in most choice situations falls outside the scope of practical reason. This in turn has upshots for our understanding of paradigmatic human agency: instead of being Enlightenment creatures who act according to the dictates of reason, we lead our lives without the guidance of reason.[17]

Pluralists can defend themselves against allegations of the irrationality of choice in several ways. First of all, they can refer to "practical wisdom"—*phronesis*—understood (p. 202) as the ability to judge that wise and virtuous people have.[18] Second, they can anticipate a "higher-level category" or a "super-value."[19] Finally, a pluralist could argue, like Ross, that rationality is based on particular intuitions that we have each time we decide, in a new situation of conflicting values, which value to choose.

Phronesis itself is a difficult notion to define. It provides, at best, rather vague guidance. Appealing to a higher level category sounds promising at first. Stocker says there may be many such categories and as an example he presents "a good day" or "a good life." If, however, "a good life" is to become/be treated as a super value, then why should it not be treated as an intrinsic value and all of the pluralist values as instrumental to this super one? Additionally, we will need to define what "a good day" or a "good life" is. If the pluralist treats everything as coming under one category, then the pluralist becomes a monist; but if there is more than one ultimate category, then the same problems of comparing values of different categories reappear. In addition, the categories are in need of further definition.

None of the discussion about the problems that confront pluralists are decisive arguments for a monistic theory of value. What we can say, however, is that monistic theories of value are simpler and face fewer problems than pluralistic theories. Particularly, monistic theories seem to be better at action guiding than pluralistic theories that have a severe problem with incomparability of values they present.

2.2. Should a Consequentialist Be a Welfarist or Nonwelfarist?

If you have decided whether you lean toward value monism or value pluralism, you can now determine whether you are interested in "how it would be best for the world to go" or "what would be best for particular people."[20] Are you concerned with *good simpliciter*, that is, good irrespective of whether it is good for anyone? Or do you believe that the good can only consist in promoting the welfare of some being or beings (leaving aside the separate question whether we limit the scope to some subgroup of humans, to all humans, or expand the circle into all sentient beings)?

(p. 203) Again, all classical utilitarians were welfarists—they believed that what we should care about is someone's welfare or well-being. Nothing is good, they stated, if it is not *good for* someone. Pleasure seemed to fulfil this condition perfectly. G. E. Moore, on the other hand, was a nonwelfarist believing that there is intrinsic value in the existence of beautiful things.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Although the division between good simpliciter and good for someone can be found in all the literature on value theories or the notion of good, it is not clear what it really conveys. Henry Sidgwick is often presented as providing the basis for this division. He famously differentiated "ultimate good on the whole for me" and "ultimate good on the whole."[21] But Sidgwick introduced these two notions, not because they were different values (once again: according to him, there was only one ultimate value—pleasure) but because the same value was differently distributed: in the first case, partially (or we may say, egoistically), in the second, impartially (some would say, ethically). Most consequentialists are impartial, perhaps even excessively so, but this does not make them either welfarists or nonwelfarists. The distinction between what is good simpliciter and what is good for someone should not be confused with the issue of how goods are distributed.[22]

Nor should the distinction be associated with the discussion whether we are capable of reaching "the point of view of the universe," that is, an objective stance that is valid for everyone, and hence can lead us to objective truths in ethics.[23] Bernard Williams and Philippa Foot[24] denied that we can know, from some kind of objective or neutral perspective, what is good for us, because they held that there is no such a thing as an objective good. Instead, they argued, what we take to be good depends on who we are, what we desire, and what projects we have. But the question of whether there is objective truth in axiology needs to be separated from the question what the substance of the (possibly objective) good for us might be. We could, for example, hold that it is an objective truth that the best for each of us would be what each of us subjectively desires.

Finally, the division into "good simpliciter" and "good for" has nothing to do with the ontological question of how values exist. Idealists[25] about values state that values exist in so far as there are people who think of them, favor them, or have some kind of positive attitude toward them.[26] Nonidealists (extreme realists) believe that whether something is good or bad is not dependent on the existence of any mind capable of making moral judgments. Goodness and badness exist like gravity. Just as the laws of physics always exist, but gravity only comes into effect when there are material objects, so values always (p. 204) exist, but only come into effect when there are sentient beings. Neither idealists nor realists are bound to accept any particular theory of the good. There can be welfarists and nonwelfarists in both groups.

If whether you are a welfarist or not says nothing about the way you want to promote the good—partially or impartially—and nothing about the epistemic probability of acquiring true value propositions, nor anything about the ontological question of whether values exist only in so far as beings who are capable of experiencing them exist, what is the purpose of presenting the distinction? What constitutes it?

It may be easier to say something positive about the distinction if we start from the other end—instead of a formal discussion, we shall turn to common language and search for our basic intuitions. What do we mean when we say that we care for our own or somebody else's well-being? Fred Feldman helps us answer the question with an image of happy parents standing over a crib of their new born babe.[27] They wish their child a good

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

life. What they wish is her well-being. Feldman admits that there is a potential problem in case a parent was irrational and had a wish that would be against the child's future will (Feldman presents an example of a religious fanatic who wants his child to become a martyr of the Lord). After all, it often happens that our parents want our lives to be different from the ones we desire to have; they do not need to want us to become martyrs!

Consequently, we should perhaps think that what is good for us is not what our parents wish for us but rather what a rational and caring person believes would benefit us. This is an understanding of well-being proposed by Stephen Darwall.[28] He retains the idea of the third-person perspective as he believes we may wish for ourselves something that is not good for us (when, for example, we want to sacrifice our own life to save someone else). But this perspective has an important objective feature: it is not defined by the desires of the subject of the life, but rather by a rational desire of this third party. In this way, Darwall solves Feldman's problem of a fanatic parent but has to face two other problems. First, it is not clear what theory of rationality he would like to accept in order to determine which benefits are rationally to be desired for someone for whom we care, and which are not. Second, it may happen that a rational and caring person will make a judgment about what would benefit us against all our wishes and beliefs.[29] From Darwall's perspective, our well-being does not depend on our own judgment, desire, or contentment at all.[30]

For us, beings who reason, have desires, and can reason upon having those desires, it may be hard to accept that our own well-being has no connection with what we want or believe. Peter Railton thinks this situation is unacceptable. In order for something to be good for someone, it "must have a connection with what he would find in some degree (p. 205) compelling or attractive, at least if he were rational and aware," explains Railton, who calls this our "internal resonance."[31] Nothing can be good for us if it does not resonate with us in some way, if it does not fulfill a resonance requirement.

> an individual's good consists in what he would want himself to want, or to pursue, were he to contemplate his present situation from a standpoint fully and vividly informed about himself and his circumstances, and entirely free of cognitive error or lapses of instrumental rationality.[32]

What is this "resonance"? Railton talks about a volitional state—desiring or wanting. We do not need to assume, however, that this exhausts the concept. Resonance can also be understood as a form of affect that manifests itself through feelings of acceptance such as liking, loving, or admiring. The feeling can be treated as a separate form of resonance, because we do not always want things that we like, and sometimes we like things we do not want. Finally, it can have a cognitive dimension—we reason and try to make a decision upon what kinds of things we would benefit from. With a cognitive dimension, it is enough to believe that a certain thing is good for us in order for it to fulfill a resonance requirement. Under such an understanding of well-being, whether the state of resonance is volitional, affective, or cognitive, what is good for us depends on our attitudes.[33]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The concept of resonance as a constitutive element of well-being is highly plausible. The plausibility comes from our common belief that we want to and should have an influence on our lives and on how well they go. We consider ourselves to be the ultimate judges of our lives, and the final authority on whether they have gone well or badly for us; or if we admit the possibility of being mistaken on that question, then at the very least, we want our opinion on that matter to be an important component of the ultimate verdict. We should not confuse the question of whether someone's life went well or badly for him with the question of what we, or others, or the judgment of history will say about that person's life. How well Stalin's life went for him is not affected by the collapse of the Soviet Union or the universal condemnation of his crimes.

Moreover, the concept of resonance is broad enough to be compatible with an almost infinite range of choices of potential intrinsic values. If you want or see value in such things as justice, freedom, or truthfulness for you, then you have a positive attitude toward it and the resonance requirement does not exclude them from being a component of what is good for you. Let us call this concept of well-being *wide-spectrum welfarism.*

Some philosophers have suggested potential problems.[34] The biggest worry is that our present attitude may suggest that something is good for us that we would regret later (having too much birthday cake). Or vice versa—we may have a negative attitude now (p. 206) toward something (going to the dentist) that in the future would prove to be beneficial for us (less pain). Which attitudes are the correct ones then? There are also other questions. How long do we need to have a positive attitude toward something to decide that it is good for us? Can we change our attitude and the good? It is easy to imagine that for a year you have a positive attitude toward, let us say, freedom but then you change your mind and you have strong affect toward justice. Which of your attitudes points to the right intrinsic value?

Railton himself could point out that he did talk of a resonance experienced by a being "fully and vividly informed about himself and his circumstances, and entirely free of cognitive error or lapses of instrumental rationality." That, at least in theory, can eliminate the aforementioned problems. As we shall see in the next section, however, it brings down other troubles instead, which do not make it easier to accept the resonance requirement.

But there may be an easier solution to these problems as well. The resonance requirement is a necessary but not a sufficient element of well-being.[35] In other words, having a positive attitude to something is *not* enough to make it intrinsically good for us. In order to decide which values are ultimate for us, we need to have a substantive discussion about values (which we will have in the next section). Still if we want to maintain a distinction between welfarism and nonwelfarism the only reasonable criterion for it is the presence or absence of resonance with the being whose welfare it is purported to be. A welfarist view, I am suggesting, holds that to be intrinsically good, something must resonate with us at least some time. On the other hand, a nonwelfarist view holds that something can be intrinsically good even if it does not resonate any time with any being.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Before we proceed to the next part, we still need to say a few words about good simpliciter or a nonwelfare good. Some thinkers hold that there are things of intrinsic value even if they do not increase anyone's well-being. They will retain this view even if we offer them our wide account of welfarism that allows them to treat as part of well-being such values as freedom, justice, or truthfulness, at the cost of acknowledging that those values are only values if they are met with a positive attitude from those who experience them. According to those who defend this view, there are simply good and bad states of affairs in the world, no matter whether anyone likes them or cares for them.[36] Moore's idea from *Principia Ethica* about a beautiful world is a classic example of such thinking.

We can find a modern example of such thinking in the school of environmentalism known as "deep ecology," founded by the Norwegian philosopher Arne Naess, which holds that:

> **1.** The well-being and flourishing of human and nonhuman life on Earth have value in themselves (synonyms: inherent worth, intrinsic value, inherent value). These values are independent of the usefulness of the nonhuman world for human purposes. (p. 207) **2.** Richness and diversity of life forms contribute to the realization of these values and are also values in themselves.[37]

For deep ecologists, life forms (including nonsentient life forms) have intrinsic value, independently of their contribution to the welfare of any sentient beings. Many of us share the intuition that it is wrong to lose a species, whether of animals or plants, that has occupied the planet we have shared for a long time. First, we feel sorry for animals that once enjoyed their lives and now die due to the changes we introduce to the wilderness. Second, in our response to natural world devastation, we also refer to our feeling of regret that the future generations, including our children, will not be able to experience "the world as it used to be," just as we are sorry we can no longer see dodo birds or dinosaurs. Third, we also fear that a fragile eco-system will collapse and make it harder or impossible for us to live on the planet. In addition, there is a feeling of guilt that we are the cause of disappearance of life, beauty, and variety in the world. But is there something more than the harm to particular animals or the usefulness of eco-systems that make us believe in the importance of the preservation of the Earth's environment? What is of value in a preserved state of the world beside the instrumental values just mentioned? If there was nobody who could appreciate or use plants, rivers, and rocks, nor has any form of consciousness that would allow them to enjoy life in any form, would it still be valuable in preserving it as it has been?

With the visible harm we are doing to other species, the deforestation and pollution we are causing, and the regret that what was once in front of our eyes is no longer there, it is difficult not to yield to the pull of our intuition that there is an intrinsic value in biodiversity. Would we still believe it, however, if we imagined a planet somewhere in the universe on which there was an enormous diversity of invisible bacteria floating in water? Perhaps we think that where there are bacteria and water, one day there will develop more complex life forms. Who knows, maybe an intelligent alien who one day will explore the universe? But if nothing, absolutely nothing more developed out of it, and there would never

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

be anyone who would in any way use the surrounding world for their flourishing in more complex ways than bacteria do, would we still believe that it is worth preserving and that it would be a great loss if there was a collision with a meteor that destroyed all the billions of bacteria? We may regret the loss of the biodiversity that existed at the time of dinosaurs (to the point that some of us would like to clone one or two of them), but knowing that we would never have developed if dinosaurs had survived may ease the pain and change our opinion on the intrinsic value of biodiversity.

2.3. Which Value?

So far, we have discussed formal problems that a consequentialist will come across while choosing a theory of value she wants to adopt. First, she needs to decide upon a number (p. 208) of intrinsic goods to promote. This may be a difficult choice between a commonsense intuition that there are a few different and irreducible intrinsic values in our lives and the philosophically simpler idea that all things that we treat as good come down to one, final intrinsic-in-itself good. The second choice a consequentialist needs to take, between *wide-spectrum welfarism* and nonwelfarism, seems a bit easier. After all, if you are a consequentialist, the idea that you promote only those goods that are welcomed, appreciated, or useful in some way to any sentient beings should be appealing. After careful reflection, Moore changed his mind about the value of beauty and in his subsequent book *Ethics* he argued that nothing can be intrinsically good unless it has some relation to a state of consciousness.[38]

But even if we reject nonwelfarism and accept wide-spectrum welfarism, we will be left with an unlimited number of values to discuss. Fortunately, we can put them into some well-known groups to make things easier for us systematically. Usually, we place theories of well-being into one of three categories: hedonism, desire-based theories, and “list theories.”

A desire-based theory is often presented as the one that fulfils the resonance requirement most straightforwardly. The intrinsic good is defined here simply in terms of what we desire—if you desire something, surely you have a positive attitude toward it. It may turn out, however, that the fulfilment of the resonance requirement is not as certain as it seems at first. It is only the simplest form of desire-based theory—one that says that your good consists in what is happening being in accord with your actual present desires—that resonates in a necessary way. According to this view, whether something is good for you is indicated by your present desire toward this thing.

Though this theory fulfils the resonance requirement, it has some unacceptable consequences. On this simple desire-based theory, something is good for you as long as you desire it, even if your desire is based on a false belief and the satisfaction of your desire would be catastrophic for you. For example, you may desire to drink what is in the glass in front of you, believing it to be a refreshing lemonade. In fact, it is a deadly poison. Is it good for you to drink it? Obviously not.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

This objection has led philosophers to produce more sophisticated desire-based theories of well-being, in which a person's actual desire is replaced by the desire that she would have if she were fully informed and thinking clearly about what she wanted.[39] When such a hypothetical desire is substituted for a person's actual desire, however, the resonance requirement is no longer met. For the actual person may never be in the required state of possessing full information and thinking clearly. If she is not, then getting what she would want if she were in that state will not give her what she actually does want, and so it may not resonate with her in any way.

A further question may be raised about this more sophisticated desire-based theory. Even if a person really is fully informed and thinking clearly, with a perfect forecast of all the consequences of satisfying his desire, is that satisfaction really good for him? Parfit's (p. 209) person with "Future Tuesday Indifference" is a counterexample to this modified desire theory.[40] This unusual, but conceivable, person doesn't care about anything that happens to him on a *future* Tuesday, so sometimes he chooses to avoid having a mildly unpleasant experience on Monday, knowing that this will mean he suffers agony on the following day. Once Tuesday is present rather than future and he is experiencing agony, he bitterly regrets his earlier choice. Yet the same thing happens week after week, because he remains indifferent to what happens to him on future Tuesdays. As this example shows, even under the condition of full information, we may be irrational and have desires that will be bad for us. But when we put a constraint on our desires and say that they should be of a rational kind, we introduce an independent standard that goes beyond resonance and toward something that has not so much connection with our positive attitudes but rather with some objective qualities. Maybe we will require people to have rational desires, but no one will want to have them.

There are also other severe problems with desire-based theories. It is not obvious how long we need to want something in order to count our desire as a "proper desire for satisfying." Let us imagine that a friend of ours wanted, for seventy years of her life, to have a secular funeral ceremony. On her death bed she changes her mind and asks to be buried in the Christian tradition. Satisfaction of which desire would be good for her: the one she had for most of her life or the one that was her "last wish"? The problem goes even further: could anything be good for her after her death? How can the satisfaction of the desire resonate with the desirer after her death when she no longer desires, feels, or believes in anything? [41]

These and other objections make it easier for some philosophers to accept a list theory. A list theory proposes a set, or list, of values that it understands as intrinsic (the "list" could contain only one item, e.g., justice). The name was introduced by Parfit, who talked of "Objective List Theory." According to such theories, "certain things are good or bad for people, whether or not these people would want to have the good things, or to avoid the bad things."[42] We can understand Parfit as saying that there are good or bad things for people even if they do not resonate with them. In this way an Objective List Theory of welfare is identical with what we have taken, in the previous section, to be a nonwelfarist theory. We then suggested that it seems reasonable to limit the way we think of well-be-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ing to the goods that resonate with us. To make things more acceptable, we offered a wide spectrum of resonance, to include desires, feelings, and beliefs. This leads us here to come up with what we will call a Subjective List Theory.[43] The goods included by a Subjective List Theory resonate with someone if that person has the appropriate evaluative beliefs about them. Suppose that the list consists of the following: justice, truthfulness, happiness, freedom, and beauty. Then supporters of this list theory (p. 210) will believe that these, and only these, values are intrinsically good for them and therefore they want to promote them in their lives.[44] This could be seen as a tempting alternative. There is a common belief among many, perhaps most people, that some set of values, if not exactly those just mentioned, constitutes our well-being. Most of us are not indifferent to them and therefore the resonance requirement is satisfied.

We should resist this tempting approach. It comes too close to a desire-based theory. Although in case of Subjective List Theory instead of *wanting* something, I *believe* it is good for me, results and problems may be similar or even identical. We may change our beliefs just as we change our desires. We might believe that getting revenge on an enemy would be good for us but in due course it may turn out that our belief was mistaken and now we believe it was bad for us to take revenge. This may push some to go back to Objective List Theory. We may trust that our lives will turn out on the whole better if we do not follow our desires or unstable beliefs but rather a "verified" set of values. Note, however, that if you accept a set of values and incorporate them as your own, you are no longer indifferent to them and they are not so much "objective" anymore; they become a subjective set.

The fact that the good for us is "subjective" in nature in such a way that its content depends on having a positive attitude toward it does not exclude the possibility of us making objective value judgments as a response to the question: "What is good for us?" It is important to clarify here two ways in which a value judgment may be objective or subjective. The move from an objective set of values to a subjective one shown in the last paragraph comes from the fact that, first, you were indifferent to certain values and then you developed a positive attitude toward them. We have said before that our own attitude is a necessary condition for something to be intrinsically good for us; this attitude is subjective in nature. The acceptance of its necessity can be stated objectively though.

It seems that neither Desire-Based Theory nor Subjective List Theory is able to provide a necessary connection between the value it offers and our resonance toward this value. Only in some, often hard-to-predict cases, is it true that satisfaction of the desire will give rise to a positive attitude in us. Similarly, with things that we believe may be good for us —we may change our attitude and the goods will turn out not to resonate with us. But what about hedonism? I will spend the rest of the paper trying to convince you that hedonism takes as the only intrinsic good a value that you not only cannot be indifferent to but also find by direct acquaintance to be intrinsically good.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

3. How Does Pleasure Resonate?

Whether and in what way pleasure resonates depends on how we define pleasure. We do not seem to have any problems with enumerating pleasures that we experience every day: good food, jogging on a sunny morning, sex with someone we love, deep (p. 211) philosophical deliberation, favorite music, and so on. We know that these are pleasures, but we want to go beyond our intuitions and learn what the quality is that unites these different things. What allows us to differentiate pleasures from other experiences?

Internalists[45] believe that what experiences of pleasure share, despite their various sources and causes, is the same "necessary quality," which determines their hedonic status. This quality is often called "hedonic tone" or simply "pleasantness" and is characteristic for both sensory experiences, such as pleasure of eating, sex, or sports, and mental or intellectual pleasures of philosophizing, reading a good book, or being in good company. According to internalists, we can experience pleasure without any positive attitude toward it. This does not mean that in practice we do not react to such an experience with a certain attitude—for example, we may and often do want to make the pleasure to last. But our attitude is not a necessary element of the pleasure experienced; it does not constitute it or cause it. At least in theory, we do not need to desire or like the state in which we are.

Externalists,[46] on the other hand, cannot see how experiences as different as sex and philosophical deliberation can have anything in common apart from our positive attitude toward it.[47] Pleasures, they say, have different sources and differ in their intrinsic nature. Externalists believe that pleasure is a kind of experience that we want for its own sake, and which we like or want to continue. In other words, it would not be pleasure if we did not want it. It is easy to see that on the externalist account, for something to be pleasure, it must, by definition, resonate with the being experiencing it.

There seems to be a deadlock between internalists and externalists who exchange convincing/persuasive arguments on both sides. In order to make a step forward, it may be useful to look for help in science. We should try to build our knowledge from two sides: on the one hand, we can refer to empirical conclusions, with the proviso that they can be incomplete and can change, and on the other hand, we can engage in careful philosophical reasoning that would help to understand the reality that scientists are discovering. We should strive for the coherence of science and philosophy, for the benefit of both. The issue of pleasure gives us a perfect opportunity to create such coherence.

Science can help us understand what kind of experience pleasure is and what its role may be. The majority of philosophers do not see, for example, a difference between (p. 212) calling pleasure a feeling and a sensation.[48] They almost universally mix the two kinds of experience as they were one.[49]

A helpful hand in understanding what pleasure is was given by an American psychologist, Magda Arnold. She starts with explaining what feeling is. According to her classification of psychic phenomena, it is "a positive or negative reaction to some experience." The positive reaction is pleasure, defined as "a welcoming of something sensed that is appraised

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

as beneficial and indicates enhanced functioning." Analogously, the negative reaction is pain, which can be understood as "a resistance to something sensed that is appraised as harmful and indicates impaired functioning." "What is pleasant is liked, what is unpleasant, disliked," she concludes.[50]

Sensations are experiences that are associated with our senses: taste, smell, sight, and so on. Sense experience "informs" what there is in the world that surrounds us, while the feeling "evaluates" what exists and signals how this affects us.[51] Only such a distinction explains why we can come through the same sensory experience in different ways: sometimes feeling pleasure, sometimes being in a neutral state of mind, and sometimes even feeling pain (depending on our attitude, mental condition, and other circumstances).[52]

Sensations are not the only experiences to which we respond with feelings of pleasure or pain; there are also thinking, imagining, or understanding.[53] This is why we usually talk of "bodily pleasures," "emotional pleasures," "intellectual pleasures," and so on. These different names that we use, strictly speaking, do not refer to pleasures but to pleasurable experiences (experience + pleasure). Experiences are different, while pleasure is of the same kind; it differs only in intensity.

Arnold's conclusions are supported by modern neuroscientists. Kent Berridge argues that "Pleasure is an additional niceness gloss painted upon the sensation."[54] Peter Shizgal repeats Arnold's assumption that there are different purposes for the sensory and hedonic systems. The first one is "to provide facts about the world"; the second gives "a subjective commentary on the information provided to them by sensory systems."[55] Nico Frijda refers to Anna Wierzbicka's linguistic research that shows that the word *pleasure*, unlike words like *good* and *bad*, cannot be found in every human language and (p. 213) therefore Frijda suggests that we should understand the word *pleasure* as "feeling good" or "the experiential property that some object feels good."[56]

We do not have the space to go further into scientific research on our topic, but this quick look seems enough to show that the conflict between internalism and externalism is only apparent. Science can help us combine the two philosophically important features. Pleasure is a feeling that arises from sensations, thoughts, images, and any other experiences, and which contains an intrinsic element of positive evaluation of these sensations, thoughts, imaginations, and other experiences. This "evaluation element" may be expressed by the acceptance of the experience, by the desire to prolong it and/or to repeat it, and, to put it simply, by recognizing that this experience is worth having in itself.

No philosophical definition is closer to the scientific explanation than that of Sidgwick. He understood pleasure as "feeling which the sentient individual at the time of feeling it implicitly or explicitly apprehends to be desirable; – desirable, that is, when considered merely as feeling, and not in respect of its objective conditions or consequences, or of any facts that come directly within the cognizance and judgment of others besides the sentient individual."[57] His remark that we should look at the feeling alone is crucial. If you believe that there are situations in which you do not like the pleasure you experience, it is because you mix the feeling of pleasure with other experiences that you are having at the

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

same time. A classic example is the feeling of guilt when having a complex experience of taking pleasure in something illegal or immoral. In such a situation, you still feel pleasure but you also feel uneasy about your experience as you know you should not be involved in it.

Thus pleasure, understood as a feeling whose nature includes the acceptance of the experience itself, liking it or treating it as something good for us, necessarily resonates with us. Though a simple desire-based theory retains resonance, it is, as we have seen, unsatisfactory in other ways, whereas the more sophisticated versions of desire-based theory no longer retain resonance. You may no longer want or like the experience that comes from the satisfaction of your desires and none of the list values seem to have the advantages of pleasure. A list theory value has this problem as well. It is possible that one day you will find one of the values good, but the next day you will change your mind. In contrast, you cannot be indifferent to pleasure, for if you are, it is not pleasure. It is by definition a feeling that, while experiencing it, you find good in itself and worthy of continuing.

4. Why Not Hedonism?

Though hedonists are lucky to have such a perfectly stable source of resonance on their side, we all know very well (at least) two strong objections that make it hard to accept (p. 214) hedonism. The first one, presented already by the Greeks, criticizes hedonists for equating all pleasures in terms of their moral value, regardless of whether they are the result of satisfying the simplest “animal” desires, such as food or sex, or the result, for example, of intellectual work.[58] For this reason Thomas Carlyle called utilitarianism the “pig philosophy.”[59] The criticism seemed so damaging that it led Mill[60] to come up with his dubious idea that there are higher and lower pleasures. The second one is Robert Nozick’s[61] “experience machine” objection. Pleasure cannot be the intrinsic good, he said, as no one would want to be plugged into a machine that would give us even the best possible states of consciousness; we want to live in reality and face it with all its pros and cons.

In response to the first objection, we may say that the problem seems less important when you understand that the feeling of pleasure should be evaluated separately from its source. It seems reasonable to accept that there is no qualitative difference in different pleasures. Your hesitation that goes: “but they feel different!” (like reading a great book and having a big portion of your favorite ice cream) comes from the fact that pleasure has different sources. Reading and eating are two different experiences, and they “feel different” no matter whether your activity is accompanied by pleasure or not. Pleasure, which makes these activities desirable for you, is still the same. You may then say that the sources of pleasure have greater or lesser instrumental value but not pleasure itself. We tend to evaluate the pleasure on the basis of the instrumental value of the activity (e.g., lying on a couch the whole day surfing the internet versus getting involved in a sophisticated philosophical discussion) and this is a mistake—if we focus on the value of the pleasure itself, it does not vary with the instrumental value of the activity that gives rise to it.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

As for the experience machine objection, for some time now it has not been treated as so serious as it used to be.[62] De Brigard[63] and Weijers[64] point out that our negative response to Nozick's imagined machine may be the result of a status quo bias, which can work the same in the opposite way—we would also be reluctant to change our current status if someone told us that we are already in an experience machine and we can be unplugged and return to reality. Silverstein[65] believes that the example is made in such a way that it reveals what we think we want, rather than what would be good for us to (p. 215) have. After all, as we have said before, our desires may be irrational or simply contrary to our own well-being.

5. Conclusion

What should a consequentialist promote? With as vast a topic as this one and as little space as we have here, I have been able to mention only a few problems and some possible solutions to them. I have attempted to show that classical hedonistic utilitarianism, as developed in its most sophisticated form by Sidgwick, offers a plausible answer to our initial question. It is what a consequentialist should promote. If we accept Railton's resonance requirement for something to count as an element of a being's welfare, and then bring a contemporary scientific understanding of pleasure to Sidgwick's philosophical understanding of it as a state of consciousness that is intrinsically worth having, we end up with something that is necessarily and intrinsically good for every sentient being experiencing it. This does not, of course, show that there are no other intrinsic values that a consequentialist should promote, but it does indicate that pleasure is unique among values in having this necessary connection with what is good for sentient beings.

References

Alston, W. P. 1967. "Pleasure." In *The Encyclopedia of Philosophy*, edited by Paul Edwards, 341–347. New York: Macmillan.

Aristotle. 2014. *Nicomachean Ethics*. Cambridge: Cambridge University Press.

Arnold, M. 1960. *Emotion and Personality. Volume I: Psychological Aspects*. New York: Columbia University Press.

Aydede, M. 2000. "An Analysis of Pleasure Vis-a-Vis Pain." *Philosophy and Phenomenological Research* LXI, no. 3: 537–570.

Berridge, K. 2010. "Fundamental Pleasure Questions." In *Pleasures of the Brain*, edited by M. Kringelbach and K. Berridge, 7–23. Oxford: Oxford University Press.

Bramble, B. 2013. "The Distinctive Feeling Theory of Pleasure." *Philosophical Studies* 162, no. 2: 201–217.

Bramble, B. 2016. "A New Defense of Hedonism about Well-being." *Ergo, An Open Access Journal of Philosophy* 3, no. 4: 85–112.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Brandt, R. 1959. *Ethical Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Carlyle, T. 1850. *Latter-Day Pamphlets*. London: Chapman & Hall.

Chang, R. 2015. "Value Incomparability and Incommensurability." In *The Oxford Handbook of Value Theory*, edited by I. Hirose and J. Olsen. Oxford: Oxford University Press.

Crisp, R. 2006. "Hedonism Reconsidered." *Philosophy and Phenomenological Research* 73, no. 3: 619–645.

Darwall, S. 2002. *Welfare and Rational Care*. Princeton, NJ: Princeton University Press.

De Brigard, F. 2010. "If You Like It, Does It Matter If It's Real? *Philosophical Psychology* 23, no. 1: 43–57.

(p. 216) Feldman, F. 2002. "The Good Life: A Defense of Attitudinal Hedonism." *Philosophy and Phenomenological Research* 65:604–628.

Feldman, F. 2004. *Pleasure and the Good Life*. Oxford: Oxford University Press.

Feldman, F. 2006. "What Is the Rational Care Theory of Welfare?" *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 3, no. 130: 585–601.

Feldman, F. 2010. *What Is This Thing Called Happiness?* Oxford: Oxford University Press.

Fletcher, G. 2016. *The Philosophy of Well-being*. Oxford: Routledge.

Foot, P. 1985. "Utilitarianism and Virtues." *Mind* 94:196–209.

Frankena, W. 1973. *Ethics*. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall.

Frijda, N. 2010. "On the Nature and Function of Pleasure." In *Pleasures of the Brain*, edited by M. Kringelbach and K. Berridge, 99–112. Oxford: Oxford University Press.

Griffin, J. 1986. *Well-being*. Oxford: Clarendon Press.

Hawkins, J. 2014. "Well-being, Time and Dementia." *Ethics* 124, no. 3: 507–542.

Haybron, D. 2008. *The Pursuit of Unhappiness*. Oxford: Oxford University Press.

Heathwood, C. 2006. "Desire Satisfactionism and Hedonism." *Philosophical Studies* 128:539–563.

Heathwood, C. 2015. "Monism and Pluralism about Value." In *The Oxford Handbook of Value Theory*, edited by I. Hirose and J. Olsen. Oxford: Oxford University Press.

Hurka, T. 1996. "Monism, Pluralism, and Rational Regret." *Ethics* 106, no. 3: 555–575.

Kagan, S. 1989. *Normative Ethics*. Boulder, CO: Westview Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Kringelbach, M., and Berridge, K., eds. 2010. *Pleasures of the Brain*. Oxford: Oxford University Press.

Labukt, I. 2012. “Hedonic Tone and the Heterogeneity of Pleasure.” *Utilitas* 24, no. 2: 172–199.

De Lazari-Radek, K. 2018. “On the Notion of Well-being. *Analiza i Egzystencja* 43:5–22.

De Lazari-Radek, K., and Singer, P. 2014. *The Point of View of the Universe*. Oxford: Oxford University Press.

Lin, E. 2016. “The Subjective List Theory of Well-being.” *Australasian Journal of Philosophy* 94, no. 1: 99–114.

Mason, E. 2018. “Value Pluralism.” In *The Stanford Encyclopedia of Philosophy* (Spring 2008 edition), edited by Edward N. Zalta. **https://plato.stanford.edu/archives/spr2018/entries/value-pluralism/**.

Mill, J. S. 1879. *Utilitarianism*. London: Longman, Green and Co.

Moore, G. E. 1903. *Principia Ethica*. Mineola, MN: Courier Corporation.

Nagel, T. 1979. *Mortal Questions*. Cambridge: Cambridge University Press.

Nozick, R. 1974. *Anarchy, State, and Utopia*. New York: Basic Book.

Oddie, G. 2013. “Value Realism.” In *The International Encyclopedia of Ethics*, edited by H. LaFollette, 5299–5310. Malden, MA: Blackwell.

Parfit, D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

Parfit, D. 2011. *On What Matters*. Oxford: Oxford University Press.

Pettit, P. 1997. Consequentialism. In *A Companion to Ethics*, edited by P. Singer. Malden, MA: Blackwell.

Rabinowicz, W., and Ronnow-Rasmussen, T. 2015. “Value Taxonomy.” In *Handbook of Value*, edited by T. Brosch and D. Sander, 23–42. Oxford: Oxford University Press.

Railton, P. 2003. *Facts, Values, and Norms*. Cambridge: Cambridge University Press.

Scanlon, T. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Schroeder, M. 2016. “Value Theory.” In *The Stanford Encyclopedia of Philosophy* (Fall 2016), edited by Edward N. Zalta. **https://plato.stanford.edu/archives/fall2016/entries/value-theory/**.

Sen, A. 1980. “Plural Utility.” *Proceedings of the Aristotelian Society* 81:193–215.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(p. 217) Shaver, R. 2016. “Sidgwick on Pleasure.” *Ethics* 126, no. 4: 901–928.

Shizgal, P. 2010. “Fundamental Pleasure Questions.” In *Pleasures of the Brain*, edited by M. Kringelbach and K. Berridge, 7–23. Oxford: Oxford University Press.

Sidgwick, H. 1907. *The Methods of Ethics*. 7th ed. London: Macmillan.

Silverstein, M. 2000. “In Defence of Happiness. A Response to the Experience Machine.” *Social Theory and Practice* 26, no. 2: 279–300.

Smuts, A. 2011. “The Feels Good Theory of Pleasure.” *Philosophical Studies* 155, no. 2: 241–265.

Stocker, M. 1990. *Plural and Conflicting Values*. Oxford: Oxford University Press.

Sumner, L. W. 1996. *Welfare, Happiness and Ethics*. Oxford: Clarendon Press.

Weijers, D. 2014. “Nozick’s Experience Machine Is Dead.” *Philosophical Psychology* 27, no. 4: 513–535.

Wierzbicka. A. 1999. *Emotions across Languages and Cultures*. Cambridge: Cambridge University Press.

Williams, B. 1981. *Moral Luck*. Cambridge: Cambridge University Press.

Williams, B. 1982. “The Point of View of the Universe: Sidgwick and the Ambitions of Ethics.” *The Cambridge Review*, May 7, 183–191.

Woolf, S. 2006. “Deconstructing Welfare: Reflections on Stephen Darwall’s *Welfare and Rational Care*.” *Utilitas* 4, no. 18: 415–426.

Notes:

([1]) Philip Pettit, *Consequentialism*, in *A Companion to Ethics*, edited by P. Singer (Malden, MA: Blackwell, 1997), 230–231.

([2]) Pettit, *Consequentialism*, 237.

([3]) Ibid.

([4]) Some philosophers use, after E. G. Moore, the phrase “intrinsic good” or “good in itself” to mean that the value of an object is grounded in its “intrinsic nature.” See, e.g., W. Rabinowicz and T. Ronnow-Rasmussen, “Value Taxonomy,” in *Handbook of Value*, edited by T. Brosch and D. Sander (Oxford: Oxford University Press, 2015), 31. To describe a value which is good on its own and not as a means to some other good, they use the phrase “final good” or “good for its own sake.” I follow, however, vocabulary used by Sidgwick, and more recently by M. Schroeder (“Value Theory,” in *Stanford Encyclopedia of Philoso-*

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

phy, edited by Edward N. Zalta, https://plato.stanford.edu/archives/fall2016/entries/value-theory/) and do not discuss here the interesting problem of "intrinsic" features of good.

([5]) On whether hedonism is a monistic or rather a pluralistic value theory, see A. Sen, "Plural Utility," *Proceedings of the Aristotelian Society* 81 (1980): 193–215.

([6]) Henry Sidgwick, *The Methods of Ethics*, 7th ed. (London: Macmillan, 1907), 401.

([7]) Hare himself was a noncognitivist and so would not have spoken of "the right act" but rather as "the act that I can prescribe universally," which is therefore the only act that, in his view, I can properly say "ought" to be done.

([8]) Marshall introduced desire theory into economics. This seems relevant: https://pdfs.semanticscholar.org/181e/830ea449c0dbca67b8cf65d83b0ef1659c6e.pdf

([9]) G. E. Moore, *Principia Ethica, Principia Ethica* (Mineola, MN: Courier Corporation, 1903), chap. III, par. 50.

([10]) Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), 501. Heathwood points out that Parfit does not commit to pluralism; see C. Heathwood, "Monism and Pluralism about Value," in *The Oxford Handbook of Value Theory*, edited by I. Hirose and J. Olsen (Oxford: Oxford University Press, 2015).

([11]) William Frankena, *Ethics*, 2nd ed. (Englewood Cliffs, NJ: Prentice-Hall, 1973), 87–88.

([12]) M. Stocker, *Plural and Conflicting Values, Plural and Conflicting Values* (Oxford: Oxford University Press, 1990), 179.

([13]) It could be claimed that hedonism has the same problem because we cannot compare different kinds of pleasure, for example those that are short and intense with those that are lasting but less intense. At least in theory, however, a sentient being gains a given amount of pleasure for each moment that a pleasure lasts, and hence simple arithmetic—multiplying the pleasure per moment by the number of moments for which it is experienced—will tell us which is the greater pleasure. (This response assumes that pleasures can be measured on a cardinal scale, and that assumption can be challenged.)

([14]) Bernard Williams, *Moral Luck* (Cambridge: Cambridge University Press, 1981), 76.

([15]) Ruth Chang, "Value Incomparability and Incommensurability," in *The Oxford Handbook of Value Theory*, edited by I. Hirose and J. Olsen (Oxford: Oxford University Press, 2015).

([16]) Chang, "Value Incomparability and Incommensurability," 205.

([17]) Chang, "Value Incomparability and Incommensurability," 206.

([18]) Nagel explains: "Provided one has taken the process of practical justification as far as it will go in the course of arriving at the conflict, one may be able to proceed without further justification, but without irrationality either. What makes this possible is judgment—

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

essentially the faculty Aristotle described as practical wisdom, which reveals itself over time in individual decisions rather than in the enunciation of general principles" (Nagel, *Mortal Questions* [Cambridge: Cambridge University Press, 1979], 135).

([19]) Stocker calls it "the higher synthesizing category": "Suppose we are trying to choose between lying on a beach and discussing philosophy—or more particularly, between the pleasure of the former and the gain in understanding from the latter. To compare them we may invoke what I will call a higher-level synthesizing category. So, we may ask which will conduce to a more pleasing day, or to a day that is better spent" (Stocker, *Plural and Conflicting Values*, 172).

([20]) T. Scanlon, *What We Owe to Each Other, What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998), 79.

([21]) Sidgwick, *The Methods of Ethics*, 112.

([22]) See as well K. deLazari-Radek, "On the Notion of Well-Being," *Analiza i Egzystencja* 43 (2018): 5–22.

([23]) Sidgwick, *The Methods of Ethics*; Bernard Williams, "The Point of View of the Universe," *The Cambridge Review*, May 7 (1981), 183–191.

([24]) Philippa Foot, "Utilitarianism and Virtues," *Mind* 94 (1985): 196–209.

([25]) I take this taxonomy from G. Oddie, "Value Realism," in *The International Encyclopedia of Ethics*, edited by H. LaFollette, 5299–5310 (Malden, MA: Blackwell, 2013).

([26]) Peter Railton, *Facts, Values, and Norms* (Cambridge: Cambridge University Press, 2003).

([27]) Fred Feldman, *Pleasure and the Good Life* (Oxford: Oxford University Press, 2004), 9–10.

([28]) Stephen Darwall, *Welfare and Rational Care* (Princeton, NJ: Princeton University Press, 2002).

([29]) Woolf, "Deconstructing Welfare: Reflections on Stephen Darwall's *Welfare and Rational Care*," *Utilitas* 4, no. 18 (2006), 421–422; Fletcher, *The Philosophy of Well-being* (Oxford: Routledge, 2016), 59.

([30]) Fred Feldman, "What Is the Rational Care Theory of Welfare?" *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 3, no. 130 (2006): 585–601.

([31]) Railton, *Facts, Values, and Norms*, 47.

([32]) Railton, *Facts, Values, and Norms*, 54.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([33]) Fletcher enumerates affective, attitudinal, and volitional states in Fletcher, *The Philosophy of Well-being*, 61–62; see also Hawkins, “Well-being, Time and Dementia,” *Ethics* 124, no. 3 (2014): 507–542, who talks of “appreciation” of value.

([34]) Fletcher, *The Philosophy of Well-being*, 65–75.

([35]) Hawkins, “Well-being, Time and Dementia.”

([36]) Mind that their statement means something else than a problem whether any being who can experience the value exists or not.

([37]) “The Deep Ecology Platform,” http://www.deepecology.org/platform.htm

([38]) Moore, *Ethics*, 249–250.

([39]) J. Griffin, *Well-being* (Oxford: Clarendon Press, 1986); L.W. Sumner, *Welfare, Happiness and Ethics, Welfare, Happiness and Ethics* (Oxford: Clarendon Press, 1996).

([40]) Parfit, *Reasons and Persons* and *On What Matters*.

([41]) For more detailed discussion on why we should reject desire-based theories, see K. deLazari-Radek and P. Singer, *The Point of View of the Universe, The Point of View of the Universe* (Oxford: Oxford University, 2014), 217–222.

([42]) Parfit, *Reasons and Persons*, 499.

([43]) See E. Lin, “The Subjective List Theory of Well-being,” *Australasian Journal of Philosophy* 94, no. 1 (2016): 99–114.

([44]) They must believe that a value is good for them; otherwise it would not resonate in this special way. How this value is to be promoted to other people is another question.

([45]) Sidgwick, *The Methods of Ethics*, 94; Sumner, *Welfare, Happiness and Ethics*, 88; A. Smuts, “The Feels Good Theory of Pleasure,” *Philosophical Studies* 155, no. 2 (2001): 241–265; B. Bramble, “The Distinctive Feeling Theory of Pleasure,” *Philosophical Studies* 162, no. 2 (2013): 201–217; I. Labukt, “Hedonic Tone and the Heterogeneity of Pleasure,” *Utilitas* 24, no. 2 (2012): 172–199.

([46]) R. Brandt, *Ethical Theory* (Englewood Cliffs, NJ: Prentice-Hall, 1959); W.P. Alston, “Pleasure,” in *The Encyclopedia of Philosophy*, edited by Paul Edwards (New York: Macmillan, 1967); C. Heathwood, *The Reduction of Sensory Pleasure to Desire*; R. Shaver, “Sidgiwck on Pleasure,” *Ethics* 126, no. 4 (2016): 901–928

([47]) The notions of internalism and externalism come from Sumner, *Welfare, Happiness and Ethics*. Internalism is sometimes referred to as “a phenomenological theory” (see Bramble, “The Distinctive Feeling Theory of Pleasure”). Externalism is sometimes called “motivational theory of pleasure” (see Heathwood, *The Reduction of Sensory Pleasure to Desire*, 24).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([48]) For a detailed discussion, see M. Aydede, “An Analysis of Pleasure Vis-a-Vis Pain,” *Philosophy and Phenomenological Research* LXI, no. 3 (2000): 537–570.

([49]) Locke wrote about “sensations,” Hume about sensations and feelings, and Sidgwick about feelings. Among modern analytical philosophers Heathwood (“Desire Satisfactionism and Hedonism”) and Feldman (*Pleasure and the Good Life*) make that mistake and talk of “sensations of pleasure.”

([50]) Magna Arnold, *Emotion and Personality* (New York: Columbia University Press, 1960), 74.

([51]) Ibid.

([52]) Imagine your different reactions to your favorite dish when you are hungry, full, and stuffed.

([53]) Arnold, *Emotion and Personality*, 75.

([54]) Kent Berridge, “Fundamental Pleasure Questions,” in *Pleasures of the Brain*, edited by M. Kringelbach and K. Berridge (Oxford: Oxford University Press, 2010), 9.

([55]) Peter Shizgal, “Fundamental Pleasure Questions,” in *Pleasures of the Brain*, edited by M. Kringelbach and K. Berridge (Oxford: Oxford University Press, 2010), 9.

([56]) Nico Frijda, “On the Nature and Function of Pleasure,” in *Pleasures of the Brain*, edited by M. Kringelbach and K. Berridge (Oxford: Oxford University Press, 2010), 101.

([57]) Sidgwick, *The Methods of Ethics*, 131.

([58]) Aristotle, *Nicomachean Ethics* (Cambridge: Cambridge University Press, 2014), 1172b; Thomas Carlyle, *Latter-Day Pamphlets* (London: Chapman & Hall, 1850), 267–270.

([59]) See also R. Crisp, “Hedonism Reconsidered,” *Philosophy and Phenomenological Research* 73, no. 3 (2006): 619–645; Feldman, *The Good Life*.

([60]) JohnStuart Mill, *Utilitarianism* (London: Longman, Green and Co, 1879), 14.

([61]) Robert Nozick, *Anarchy, State, and Utopia, Anarchy, State, and Utopia* (New York: Basic Books, 1974), 42–45.

([62]) Lin, “The Subjective List Theory of Well-being.”

([63]) F. De Brigard, “If You Like It, Does It Matter If It’s Real?” *Philosophical Psychology* 23, no. 1 (2010): 43–57

([64]) D. Weijers, “Nozick’s Experience Machine Is Dead,” *Philosophical Psychology* 27, no. 4 (2014).

([65]) M. Silverstein, “In Defence of Happiness,” *Social Theory and Practice* 26, no. 2 (2010): 279–300.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Katarzyna de Lazari-Radek

Katarzyna de Lazari-Radek is Assistant Professor at the Faculty of Philosophy, University of Łódz, Poland. She is a hedonistic utilitarian. Her main research interest focuses on the philosophy of Henry Sidgwick and Derek Parfit, as well as the concept of well-being and pleasure. Together with Peter Singer she wrote two books: The Point of View of the Universe (Oxford University Press, 2014) and Utilitarianism—A Very Short Introduction (Oxford University Press, 2017). Apart from academic work, she is keen to convey philosophical ideas to a wider audience, giving lectures and writing for popular magazines on how to live a good life.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Actualism, Possibilism, and the Nature of Consequentialism

Yishai Cohen and Travis Timmerman

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.5

Abstract and Keywords

The actualism/possibilism debate in ethics is about whether counterfactuals of freedom concerning what an agent would freely do if she were in certain circumstances even partly determine that agent's obligations. This debate arose from an argument against the coherence of utilitarianism in the deontic logic literature. In this chapter, we first trace the historical origins of this debate and then examine actualism, possibilism, and securitism through the lens of consequentialism. After examining their respective benefits and drawbacks, we argue that, contrary to what has been assumed, actualism and securitism both succumb to the so-called nonratifiability problem. In making this argument, we develop this problem in detail and argue that it's a much more serious problem than has been appreciated. We conclude by arguing that an alternative view, hybridism, is independently the most plausible position and best fits with the nature of consequentialism, partly in light of avoiding the nonratifiability problem.

Keywords: consequentialism, actualism, possibilism, securitism, hybridism, maximalism, omnism, counterfactuals of freedom, action-guiding, nonratifiability problem

1. Introduction

WE will use the term *consequentialism* to pick out views in normative ethics which hold that the deontic status of an act-set is solely a function of the net value of the consequences of performing the act-set in question, relative to the net value of the other act-sets that the agent can perform.[1,2] Consequentialist views are demarcated from one another by (i) their evaluative rankings of outcomes and, more fundamentally, (ii) their account of the features that determine the evaluative rankings of outcomes. For example, with respect to (i), two consequentialist views may disagree about whether an outcome in which one billion people are prevented from experiencing a headache is better than an outcome in which one innocent person is prevented from dying. With respect to (ii), this disagreement may be the result of each view identifying different intrinsic goods and bads. Perhaps preventing one billion people from experiencing a headache maximizes he-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

donic utility, while preventing the innocent person from dying maximizes preference satisfaction.[3]

(p. 140) Category (ii) ranges over a wide variety of issues, though the literature on consequentialism has primarily focused on identifying the correct axiology (e.g., what is intrinsically good and bad), identifying the types of reasons that exist (e.g., moral, prudential), and identifying their respective weights. These distinctions do not, however, exhaust the differences between consequentialist views. Any form of consequentialism must also take a stance on the scope of possible acts that are relevant *options* for the agent given certain facts about how an agent would (or might) freely act under certain circumstances. These relatively recent issues in the history of ethics center on the nature of control that an agent must have over their behavior, both in the immediate future and across one's life. The question of the scope of an agent's options most notably arises in the actualism/possibilism (A/P) debate, which is the focus of this paper.[4]

The A/P debate may be illustrated with the following case. Suppose that you've just been invited to an ex-partner's wedding, and while you can attend the wedding and be pleasant at little cost to yourself, you're also prone to fits of jealousy. The *best* act-set you can perform involves <accepting the invitation, attending the wedding, and staying sober>, ensuring that everyone has a good time. The *worst* act-set you can perform involves <accepting the invitation, attending the wedding, and getting inebriated>, thereby ruining the wedding for everyone. The act-set you can perform that is neither the best nor the worst involves <declining the invitation, and doing something else besides attending the wedding>.[5] Finally, suppose that if you were to accept the invitation today, then you would, in fact, get inebriated once you're at the wedding.

The question at the root of the A/P debate is whether you are obligated to accept or decline the wedding invitation. This question is particularly important for consequentialists since they must take a stance on whether true counterfactuals of freedom (CFs)—true subjunctive conditionals concerning what an agent S would freely do at t_2 if S were in circumstances C at t_1—even partly determine an agent's moral obligation to perform some act-set.[6],[7]

There are four types of views in the debate, viz. possibilism, actualism, securitism, and hybridism, and each of these types admits of multiple distinct precisifications. In response to the central question of the debate, possibilists and some hybridists believe that, necessarily, true CFs do not even partly determine an agent's obligations, while actualists, securitists, and some hybridists believe that true CFs can at least partly (p. 141) determine an agent's obligations. So, with respect to the wedding invitation example, actualist forms of consequentialism maintain that you are obligated to decline the invitation at least partly because what would actually happen if you were to decline is better than what would actually happen if you were to accept. Possibilist forms of consequentialism hold that you are obligated to accept because doing so is part of the best act-set you can perform over time. Securitist and hybridist forms of consequentialism are a bit more com-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

plex. We will have to consider more detailed versions of the wedding invitation case before we are in a position to explain the implications of these views.

Our aim is to elucidate the differences between actualist, possibilist, securitist, and hybridist forms of consequentialism and, in doing so, explore the benefits and drawbacks of each from a consequentialist perspective. This chapter is structured as follows. In section 2 we provide some historical context, briefly explaining how a supposed puzzle for utilitarianism within the context of deontic logic gave rise to the actualism/possibilism literature, and then in section 3 we discuss time's crucial relationship to abilities and obligations. In sections 4–8, we review possibilist, actualist, securitist, and hybridist forms of consequentialism in more detail and argue that hybridism, in one form or other, is both the most plausible view and best captures the nature of consequentialism, partly in light of avoiding the so-called nonratifiability problem.

2. Consequentialism and Deontic Logic

The purpose of this section is to trace the historical origins of the A/P debate and explain its relation to consequentialism. The debate arose partly from Hector-Neri Castañeda's (1968) argument that utilitarianism is formally incoherent. Here's the concise version. Castañeda first assumed that utilitarians accept a principle of deontic logic known as "obligation distributes through conjunction." This principle holds that if an agent *S* is obligated to do both A and B, then *S* is obligated to do A and *S* is obligated to do B (1968, 141). This principle may be represented more formally as follows.

> **Obligation Distributes Through Conjunction:** O(A & B) → O(A) & O(B)

This principle, hereafter (ODC), can be illustrated with the wedding invitation case. If you are obligated to <accept the wedding invitation and stay sober at the wedding>, then you are obligated to <accept the wedding invitation> and you are obligated to <stay sober at the wedding>.

Second, Castañeda (1968, 142) argued that (ODC) is inconsistent with the following principle, which he took to be a basic commitment of utilitarianism.

> **(U):** *S* is morally obligated to do *x* in circumstances *C* iff *S*'s doing *x* in *C* will bring about a greater balance of good over bad than her performing any other alternative action open to her in *C*.

(p. 142) To see why (ODC) and (U) are supposedly inconsistent, suppose that an agent *S*'s performing the conjunctive act-set <A & B> brings about a greater balance of good over bad than any alternative act-set (singleton or plural) that *S* can perform. It is supposed to follow from (U) that *S* is obligated to perform <A & B>. Now, given (ODC), it follows that *S* is obligated to perform <A> and that *S* is obligated to perform <B>. But, given (U), performing <A> would result in more net good than performing any alternative, including <B>, *and* performing <B> would *also* result in more net good than performing any alter-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

native, including <A>. But performing <A> cannot result in both more and less net good than performing <B>.[8]

Dag Prawitz (1970) and Fred Westphal (1972) responded to Castañeda by modifying (U) in such a way that the actions are indexed to the time they would need to be performed in order to bring about the uniquely optimific outcome. On their suggested revision, if performing the act-set <A & B> from t_1-t_2 would result in the greatest net amount of good, then *S* is obligated to perform <A> at t_1 and to perform <B> at t_2. This supposedly avoids the contradiction because <A> supposedly is *the* act-set at t_1 that would produce the greatest net amount of good in comparison to any other act-set performable at t_1, whereas <B> supposedly is *the* act-set at t_2 that would produce the greatest net amount of good in comparison to any other act-set that is performable at t_2. Finally, from among the performable act-sets that might occur from t_1-t_2, <A & B> is *the* act-set that would produce the greatest net amount of good in comparison to any other act-set that is performable from t_1-t_2. Thus, each act-set is the uniquely optimific one at the time(s) it is performed.[9]

Harold Zellner (1972) demonstrated that indexing actions to times will not solve the supposed problem Castañeda identified for utilitarianism. While performing <A & B> from t_1-t_2 may be uniquely optimific, it does not follow that performing either of these individual acts at their respective times would be uniquely optimific.[10] For instance, performing <A> at t_1 may not be uniquely optimific if the agent would *not* perform <B> at t_2 if she were to perform <A> at t_1. This point may be illustrated with the following case.

> **Covering Class:** The best act-set Bill can perform is <A & B> from t_1-t_2, where <A> = agree to teach Ted's class next week, and <B> = teach Ted's class next week. Bill can also <C> at t_1, where <C> = suggest that George, an inferior instructor, cover Ted's class instead. Finally, suppose that if Bill were to <A> agree to teach Ted's class next week, he would <~B> not teach class.

(p. 143) Again, it would be best if Bill <agrees to teach and teaches>, second best if he <suggests that George cover the class>, and worst if he <agrees to teach and skips teaching>. Thus, the value of the act-sets may be ranked from best to worst as follows.

<A & B>

<C >

<A & ~B>

Zellner points out that, since Bill would <~B> if he were to <A>, the value of performing <A> is not uniquely optimific at t_1, even though the value of performing <A & B> from t_1-t_2 is uniquely optimific. So, in such cases, (U) combined with (ODC) still generates contradictions, even if each act-set (singleton or plural) is indexed to their respective times.

Zellner's own proposed solution was to reject (U) because it is inconsistent with a supposedly basic principle of inference which holds that if an agent is obligated to perform <A> and her performing <A> entails her performing <B>, then she is obligated to perform

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

<B> (1972, 125). This rule, sometimes referred to as *Normative Inheritance* (NI), may be represented more formally as follows (Feldman 1986, 41).

Normative Inheritance: If $\vdash A \rightarrow B$ then $\vdash$ **O(A)** $\rightarrow$ **O(B)**

Regardless of whether Zellner's own response to Castañeda succeeds, cases such as *Covering Class* highlight the important question within the A/P debate as to whether true CFs even partly determine an agent's obligations.

Since a conjunction entails each of its conjuncts, it follows that (NI) entails (ODC). Moreover, a commitment to (NI), and thus to (ODC), suggests that, in addition to an entailment relation, there is also a dependence relation between obligatory act-sets. We can distinguish between a *nondependent* obligation which we have not in virtue of having some other obligation and a *dependent* obligation which we have in virtue of having some other obligation. All dependent obligations are ultimately possessed in virtue of possessing some nondependent obligation (cf. Portmore 2011, 179; Timmerman and Cohen 2016, 679).

Many such theories that are committed to (ODC) or (NI) are forms of maximalism. They maintain that, necessarily, the object of a nondependent obligation for an agent *S* is a maximal act-set, or something close enough.[11] A maximal act-set is roughly one that, at some time *t*, *S* can perform across their entire life up to the last performable act. More precisely, a maximal act-set is an act-set that, at some time *t*, an agent *S* can perform over time, and it is not contained in some other act-set that, at *t*, *S* can perform over time. Act-set *x* is contained in act-set *y* iff (i) *x* and *y* belong to the same agent, (ii) the period (p. 144) of time at which *x* is performed is a proper or improper part of the period at which *y* is performed, and (iii) it is logically necessary that if *y* is performed then *x* is performed (Brown 2018, 752; Portmore 2011, 177; Prawitz 1968, 80; Sobel 1976, 199). The final crucial component to maximalism is that every nonmaximal act-set that is contained in a nondependent, obligatory maximal act-set is itself the object of a dependent obligation.

In contrast to maximalism, omnism implies that the object of a nondependent obligation may be a maximal or nonmaximal act-set because the deontic status of every act-set *x*, maximal and nonmaximal, is to be evaluated only in terms of its own value rather than in terms of its relation to the value of a distinct act-set *y* that entails *x* or which contains *x* (Portmore 2017, 429, 431). With the exception of the version of actualism to be discussed in section 5, all of the theories to be discussed presuppose maximalism, or something similar to maximalism.

In the next section, we provide an elaborated version of the wedding invitation case that will guide us through the A/P terrain, and we will highlight a number of background assumptions that we explicitly adopt in our approach to the debate.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

3. Time’s Relationship to Abilities and Obligations

Before we can explore the views of the A/P debate in detail, we must first discuss time’s role with respect to an agent’s abilities and obligations. We’ll do so, in part, through further discussion of the aforementioned principles of deontic logic. This will provide the background information necessary to understand the subsequent detailed discussion of each of the views in the A/P debate.

Let’s start by considering an elaborated version of the wedding invitation case, which we’ll refer to as *Wedding Invitation 1* (*WI1*). Suppose that Alice and her ex-partner promised each other in the past that they would attend each other’s weddings, and that Alice has been invited to her ex-partner’s wedding. At t_1 Alice is <A>: deliberating about what to do. The following actions are ones that, at t_1, she can perform in the future:

Possible actions at t_2:

B = Decide to: accept the wedding invitation at t_2, and then attend the wedding and not drink alcoholic beverages from t_3-t_4

C = Decide to: decline the wedding invitation at t_2, and then go home and do research for a paper from t_3-t_4

Possible actions at t_3:

D = Decide to: attend the wedding and not drink alcoholic beverages from t_3-t_4

E = Decide to: go home and play video games from t_3-t_4

(p. 145) **F** = Decide to: go home and do research for a paper from t_3-t_4

G = Decide to: go home and play video games from t_3-t_4

Possible actions at t_4:

H = Decide to: not drink alcoholic beverages at t_4

I = Decide to: drink an alcoholic beverage at t_4

J = Decide to: play video games at t_4

K = Decide to: do research for a paper at t_4

L = Decide to: play video games at t_4

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

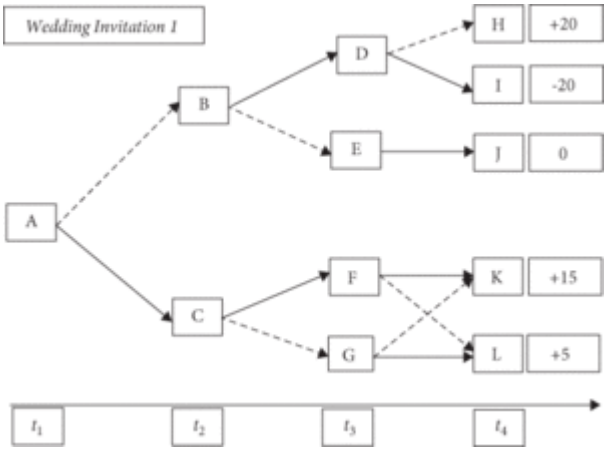


Figure 7.1. Wedding Invitation 1.

All arrows point to what Alice *can* subsequently do, but only the solid arrows indicate what Alice *would* subsequently do if she were in certain prior circumstances (see Figure 7.1). Notice, then, that, no matter what Alice decides at t_2 to do from t_2-t_4 among the decisions that, at t_1, she can make at t_2, it's not the case that Alice *would* perform any of the actions (more specifically, decisions) that are preceded at any point in time by a dotted line. The positive and negative numbers represent the total net hedonic utility (i.e., the total number of hedons and dolors) of the consequences that would be brought about by each of the aforementioned act-sets. To keep things simple, let's suppose that hedonistic act-utilitarianism is true, though, to be clear, all arguments put forward in this chapter can accommodate any version of consequentialism, including those that posit agent-relative constraints and/or pluralism about intrinsic value. Let's also suppose for simplicity's sake that, for whatever reason, the final act that Alice can perform extends no later than t_4.

At t_1, <B, D, H> is the best maximal act-set that Alice can perform over time, <B, D, I> is the *worst* maximal act-set that Alice can perform over time, and <C, F, K> is the (p. 146) *second-best* maximal act-set she can perform over time. As a matter of fact, Alice performs the second-best maximal act-set that, at t_1, she can perform from t_2-t_4. Moreover, if Alice were to <B> at t_2, then she would, in fact, <D> at t_3 and <I> at t_4.

With this case in mind, let's consider what Alice's abilities are at any given time. It is natural, we think, to understand an agent's *present* ability to do something as an ability to bring about a *future* event that might occur rather than as an ability to bring about a present event that is already happening, though many have suggested that this distinction doesn't have any significant ramifications for the A/P debate.[12] We will adopt an exclusively future-oriented approach, according to which an agent has a present ability to settle which of the alternative *future* options will be realized.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

4. Possibilism

Now that we have reviewed the relevant background information, we are in a position to dive deeper into the A/P debate by considering each of the views in more detail. In this section, we will examine possibilism and explain why it is viewed as an attractive position, while also reviewing the most compelling objection against it. Here is a more formal definition of possibilism.

> **Possibilism:** At t an agent S is obligated to φ at t' iff S can, at t, φ at t' and φ-ing at t' is part of the best act-set that, at t, S can perform from t' until the last time she can perform an act.

According to possibilism, an agent's sole *nondependent obligation* is to perform an optimific *maximal act-set* (i.e., the best act-set agents can ensure they perform from the time in question to the last time they can perform an act), and they will have *dependent obligations* to perform the nonmaximal act-sets that are contained in this nondependent, obligatory maximal act-set. In *WI1*, for instance, possibilists hold that Alice has a dependent t_1 obligation to <B> at t_2 in virtue of having a nondependent t_1 obligation to <B, D, H> from t_2-t_4.[13]

Possibilism has a number of defenders for a reason. In addition to generating the intuitively correct moral verdicts in a wide range of cases, it also comports well with consequentialist commitments. Specifically, possibilism captures the most straightforward understanding of the idea that agents are obligated to bring about the best outcome they can. It's for this reason that consequentialists[14] seem to frequently implicitly (p. 147) assume possibilism in their work. A second set of advantages is that possibilism preserves a number of plausible axioms in deontic logic, most notably (ODC) and (NI).[15] A third set of advantages is that possibilism avoids the primary objections to actualism to be discussed in the next section.[16] While possibilism no doubt has a lot going for it, it is also subject to a variety of objections, one of which is particularly notable.

The most influential objection to possibilism is that it can generate an obligation that, if acted on, would result in the worst possible outcome.[17] Possibilism generates this consequence because it implies that true CFs do not even partly determine an agent's obligations. This objection may be formulated more precisely as follows.

> **Worst Outcome Objection:** Possibilism entails that an agent S can have an obligation to φ even when φ-ing would result in S performing an act-set that is deeply morally wrong (perhaps the worst possible act-set) and that is worse than the act-set that S would perform if S were to $\sim\varphi$.

To illustrate, possibilism says that in *WI1* Alice has a dependent t_1 obligation to <B, D> from t_2-t_3 in virtue of having a nondependent t_1 obligation to <B, D, H> from t_2-t_4. However, if Alice were to <B, D>, then she would <I> rather than <H> at t_4, resulting in the worst possible outcome. This may not sound so counterintuitive in cases where the worst outcome isn't tragic. Yet this objection has more intuitive force in high-stakes variants.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Suppose that, no matter what Alice decides prior to t_4 to do, if Alice were to drink alcohol at the wedding, then she would drive home drunk, killing three pedestrians in the process. Possibilism still renders the verdict that Alice has a nondependent t_1 obligation to <B, D> from t_2-t_3, and it renders this verdict no matter how terrible the consequences of doing this happen to be. So, while possibilism can account for consequentialists' judgments that agents are obligated to bring about the best outcome that they can over time, they cannot account for consequentialists' judgments that agents are always obligated to act in ways that *would* bring about the best outcome.

Possibilists have tried to sugar the pill of the *Worst Outcome Objection* by emphasizing the distinction between conditional and unconditional obligations. They'll note that the possibilist obligation picks out agents' unconditional obligations and still allows that agents can incur conditional obligations to act in ways that preclude them from bringing about the worst outcome.[18] So, while it's true that possibilism entails that Alice has an unconditional obligation to <B> in virtue of her unconditional obligation to <B, D, H>, it may also be true that she has a conditional obligation to <C> given that Alice would (p. 148) <I> if she were to <B>. More generally, possibilists respond to this objection by claiming that agents have an unconditional obligation to do the best they can but incur conditional obligations to bring about the next best outcome *if* they won't bring about the best outcome. While the appeal to conditional obligations renders possibilism more palatable, the *Worst Outcome Objection* nevertheless reveals an important tension between possibilism and consequentialist commitments.

5. Actualism

In this section, we will examine a standard form of actualism and then review two problems with actualism that concern the relationship between control and the truth value of certain CFs. Here is a standard definition of actualism.

> **Actualism:** At t an agent S has an obligation to φ at t^* iff, at t, S can φ at t* ($t < t^*$), and what would happen if S were to φ at t^* is better than what would happen if S were to perform any other act-set that, at t, S can perform at t^*.[19]

This standard definition of actualism implies that Alice has a t_1 obligation to <C> at t_2, not because it is contained in the best act-set that, at t_1, Alice can perform over time, but rather because what would happen if Alice were to <C> at t_2 is better than any other act that, at t_1, Alice can perform at t_2. Actualism thus avoids the worst outcome objection precisely by maintaining that true CFs at least partly determine an agent's obligations.

One problem with this definition is that it does not require φ-ing at t^* to be a *fully specified* act-set, that is, an act-set that is not contained in any other act-set that, at t, S can perform at t^*.[20] To see why "φ" needs to be fully specified, consider another wedding scenario, *WI2*, that has the following deviations from *WI1*.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Alice's neighbors always have loud and disruptive parties, and, rather than <K> and <L> being options for Alice at t_3, Alice can, at t_3, perform any of the following acts at t_4:

K1 = <Decide to: do research and kill the neighbors at t_4>

K2 = <Decide to: do research and *not* kill anyone at t_4>

L1 = <Decide to: play video games and kill the neighbors at t_4>

L2 = <Decide to: play video games and *not* kill anyone at t_4>

(p. 149) Suppose that, as a matter of stipulation, Alice would <L2> at t_4 if she were to <F> at t_3, and that if doing research were part of the content of Alice's decision at t_4, then killing the neighbors would also be part of the content of Alice's decision at t_4 (the closest world in which Alice does research is one in which she kills the neighbors). Nevertheless, in light of the fact that, at t_3, Alice can <K2> at t_4, we may conclude that Alice is still obligated to form a decision whose content includes doing research since <K2> is the best fully specified act-set at t_4 that, at t_3, Alice can bring about. The lesson to be gleaned from this, according to Goldman (1978, 186–190), is that no agent has an obligation even partly in virtue of the truth value of what we will call a *synchronic* CF—that is, a counterfactual in which the antecedent and the consequent are both indexed to the same time. So, the following true synchronic CF does not even partly determine Alice's obligations at t_3: "If Alice were to <do research> at t_4, then Alice would also <kill the neighbors> at t_4."[21] This is why the object of an agent's obligation in any A/P scenario must be a fully specified act-set. So, we will suppose that each of the act-sets mentioned in *WI1* is fully specified and that the content of each decision implicitly excludes the performance of any other number of normatively significant acts like killing the neighbors or donating to charity. Goldman's remarks reveal that the kind of control an agent must have in order to have a moral obligation involves a control over the truth value of the relevant synchronic counterfactuals.

Another kind of counterfactual that does not even partly determine an agent's present obligations concerns what an agent would subsequently freely do if she were in the very circumstances in which she presently finds herself. For instance, when considering what, at t_1, Alice ought to do at t_2, all parties in the debate assume that what Alice would freely do at t_2 if she were to **<A>** at t_1 does not even partly determine Alice's t_1 obligations. We will call this kind of counterfactual an *early* counterfactual because the antecedent involves a time t at which an agent has an obligation to subsequently do something at t^* ($t < t^*$). No party in the A/P debate assumes that early counterfactuals even partly determine an agent's obligations. To illustrate this point, consider a scenario, *WI3*, that is similar to *WI1*, except that Alice performs <B, D> from t_2-t_3. Actualists and possibilists alike agree that, at t_3, Alice ought to <H> at t_4, and so the following true CF does not even partly determine Alice's t_3 obligation to do something at t_4: "If Alice were to <D> at t_3, then Alice would <I> at t_4." After all, the truth of this early counterfactual is incompatible with Alice's performing <H> at t_4, and at t_3 Alice can do something at t_4, viz. <H>, such that if she were to do it, then this early counterfactual would be false (Cohen and Timmerman

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

2016). Having control over the truth value of an early counterfactual highlights the importance of an agent's control over their immediately available action(s). A more plausible position in the A/P debate must take these insights into account, and securitist views of different stripes aim to do exactly that.

(p. 150) 6. Securitism

Securitist views may be thought of as versions of actualism that took these insights into account and were modified to focus on the best outcome that would occur from among all of the fully specified act-sets that an agent can immediately perform. In this section, we will briefly review the most important kinds of securitism in the literature before reviewing two of the most influential objections that apply to each view. Perhaps the most popular kinds of securitism are maximalist versions, which hold that an agent's options at t are the jointly exhaustive and mutually exclusive, fully specified, maximal act-sets that are securable for an agent at t.[22] Here is a formal definition of maximalist securitism.

> **Securitism:** At t an agent S has a nondependent obligation to φ at t^* ($t < t^*$) iff, φ is a fully specified, maximal act-set that, is securable for S at t, and what would happen if S were to φ at t^* is better than what would happen if S were to perform any other fully specified, maximal act-set that is securable for S at t.

A maximal act-set L[23] is securable for S at time t if, at t, S can immediately perform the first moment of x, x is in L, and if x were to occur, then L would occur (Sobel 1976, 199). Not all forms of securitism are maximalist, however. Goldman's version of securitism is neither maximalist nor omnist. Instead, Goldman (1978, 194–195) maintains that, at t, an agent can (in the relevant sense of "can") perform an act-set over time only if, at t, an agent can form an intention to perform this act-set, the act-set is securable in virtue of the decision being causally efficacious if performed in the sense that the content of the decision would be actualized if the decision were to occur, and this content corresponds to the relevant securable act-set.[24] We stipulated in *WI1* that the content of the decisions involve all of the relevant activity up to t_4 in order to make Goldman's verdict about *WI1* align with the verdict of maximalist versions of securitism, such as Portmore's version. Both hold that Alice has a nondependent t_1 obligation to <C, F, K> from t_2-t_4, and that this act-set is securable for Alice at t_1 in virtue of Alice having, at t_1, the ability to perform the fully specified singleton act-set <C> at t_2.

We now turn to the most important objections in the literature that apply to both actualism and securitism. The first influential objection is that both views let agents off the hook too easily by allowing them to avoid incurring moral obligations in virtue of their rotten moral dispositions.[25] In *WI1*, for instance, both views allow Alice to avoid (p. 151) incurring an obligation to attend the wedding simply because she is disposed to act wrongly if she were to freely decide to accept the invitation. With respect to securitist views, it matters that it is not securable at t_1 for Alice to both attend the wedding and refrain from drinking from t_3-t_4 (Timmerman 2015; Vessel 2016). Nevertheless, attending the wedding at t_3 is easily securable for her at t_1, and if Alice were to attend, then not

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

drinking at t_4 would be easily securable for her at t_3.[26] But, it seems, agents shouldn't get out of having to do good things just because they're disposed to do bad things. This objection may be stated more precisely as follows.

> **The Not Demanding Enough Objection:** Actualism and securitism permit an agent *S* to avoid incurring any moral obligation to *φ*, which S can easily fulfill, simply in virtue of *S*'s rotten moral character.[27]

The second influential objection is that both views prescribe the agent to perform harmful actions even when the agent can refrain from performing such actions. To take an extreme example, in certain cases they could require someone to murder five innocent people if he would otherwise murder six innocent people, even though (everyone agrees) the agent can refrain from murdering anyone. In the less extreme example of *WI1*, both views entail that, at t_1, Alice has a dependent obligation to <C> at t_2 and that, as a result, actualism and securitism prescribe the bad behavior of breaking a promise to attend someone's wedding. That's not quite as bad as murdering five innocent people, but it still seems objectionable, given that Alice can attend the wedding and refrain from drinking. This objection may be formulated more precisely as follows.

> **The Bad Behavior Objection:** Actualism and securitism prescribe bad behavior, and acting on such prescriptions presumably renders[28] an agent *S* immune from moral criticism, even when *S* can easily refrain from such behavior.[29]

In the next section we analyze an underexplored objection, the nonratifiability problem, and argue that, contrary to what has been assumed, all versions of actualism and securitism are subject to this problem, and that this problem is more serious than has been assumed in the literature. Considering this problem will also help motivate hybridism, which we discuss in section 8.

(p. 152) 7. The Nonratifiability Problem

Securitists such as Goldman (1978, 202) and Portmore (2011, 179, 181–182), maintain that, in *WI1*, Alice has at least three dependent t_1 obligations viz. to <C> at t_2, <F> at t_3, and <K> at t_4, respectively, and that Alice has these obligations in virtue of having a nondependent t_1 obligation to <C, F, K> from t_2-t_4. Moreover, both agree that Alice has a (nondependent) t_3 obligation to <K> at t_4, which is exactly what we would expect given that Alice violates no obligation (according to both actualism and securitism) from t_2-t_3, and given that Alice has a (dependent) t_1 obligation to <K> at t_4. However, here's a problem for such views. Fulfilling one's obligations during some period of time does *not* guarantee a consistency among an agent's various dependent singleton obligations across time. To see the problem, consider a revised version of the wedding case, *WI4* (see Figure 7.2), in which Alice performs <L> rather than <K> at t_4, and so the following CF that is false in *WI1* is true in *WI4*: "If Alice were to <F> at t_3, then she would <L> at t_4." Let's also suppose that Alice has the following additional option in our new case, *WI4*:

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

M = Decide to: decline the wedding invitation at t_2, and then go home and play video games from t_3-t_4

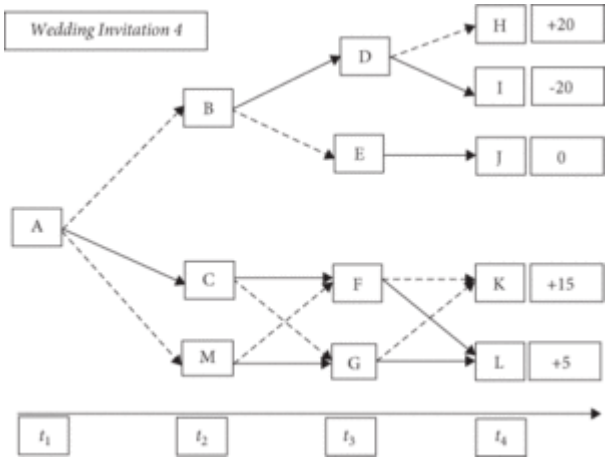


Figure 7.2. Wedding Invitation 4.

WI4 reveals that both actualism and securitism allow for a conflict among an agent's obligations across time, even when an agent never violates an obligation during that interval of time. Both views hold that Alice has a dependent t_1 obligation to <L> at t_4 in virtue of having a nondependent t_1 obligation to perform a maximal act-set that con (p. 153) tains <L>. Alice has this nondependent t_1 obligation because such an act-set is a better securable one for Alice at t_1 than any other securable act-set for Alice at t_1 that does not contain <L>, viz. <B, D, I> (+5 > –20). Nevertheless, regardless of whether Alice performs <F> or <G> at t_3, Alice has a t_3 obligation to <K> rather than <L> at t_4 (+15 > +5), even though <L>, and not <K>, is one of Alice's dependent t_1 obligations. Ross (2012, 87–89) calls this phenomenon the nonratifiability problem: actualism and securitism make nonratifiable prescriptions, that is, prescriptions that will inevitably be reversed prior to the time of the act, no matter what the agent does prior to that act.

There is a specific sense in which actualism and securitism appear *not* to be action-guiding in light of these conflicting obligations across time; it is difficult to see in what sense Alice has a t_1 obligation to <L> at t_4 since, no matter what she does prior to t_4—and, we would like to emphasize, even if she never violates an obligation prior to t_4—Alice lacks a t_3 obligation to <L> at t_4. Possibilism, by contrast, faces no such problem: just as Alice has a dependent t_1 obligation to <H> at t_4 (in virtue of having a nondependent t_1 obligation to <B, D, H> from t_2-t_4), Alice would similarly have a t_3 obligation to <H> at t_4 if Alice were to <B, D> from t_2-t_3.

We think that the nonratifiability problem raises additional unexplored difficulties for actualism and securitism. But before we explore such difficulties, we will first turn to Ross's theory and explain how it aims to avoid the nonratifiability problem. We will argue that, contrary to what Ross claims, his theory is not immune to the nonratifiability problem. Here is Ross's (2012, 91) version of securitism.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> **Momentwise Wide-scope Securitism (*MWSS*):** For any *x* and *t*, at *t*, *x* ought to be such that, for all *t'* from *t* forward, *x* satisfies the following conditional: For all *Φ*, if whether *x Φ*s does not causally depend on the intentions *x* has after *t'*, and if every maximally preferable option that is directly securable for *x* at *t'* involves *Φ*-ing, then *x Φ*s.

Like *WI4*, Ross considers a case that highlights the nonratifiability problem. Here is a summary of the case (Ross 2012, 87–88):

> *Satan's School for Girls* On June 6, 2011, Sally is about to be kidnapped by Satanists and incorporated into Satan's School for Girls. On June 6, 2016, Sally can kill her firstborn child (for the prince of darkness) either by cutting off her child's head or by bludgeoning her child to death with an axe, and Sally can also refrain from killing her child. However, while both cutting off her child's head in 2016 and killing her child with an axe in 2016 are securable for Sally in 2011, refraining from killing her child in 2016 is not securable for Sally in 2011. In other words, even if Sally were to decide in 2011 to refrain from killing her child in 2016, she would still kill her child in 2016.

The best maximal act-set that is securable for Sally on June 6, 2011, involves cutting off her child's head on June 6, 2016 (this form of death is less painful). However, at the relevant time on June 6, 2016, the best maximal act-set that is securable for Sally involves (p. 154) refraining from killing her child in 2016. Like *WI4*, we may wonder in what sense Sally has a 2011 obligation to kill her child in 2016 since, even if Sally fulfills all of her obligations up to June 6, 2016, Sally will not be obligated to follow through on her 2011 obligation to cut off her child's head in 2916. Now, Ross claims that, unlike securitism, *MWSS* avoids the nonratifiability problem:

> MWSS implies that, on June 6, 2011, Sally is under an obligation to refrain from killing her firstborn child five years later. For MWSS implies that, on June 6, 2011, Sally ought to be such that, *on June 6, 2016*, she satisfies the conditional (if whether Sally refrains from killing her firstborn child on June 6, 2016 does not depend on her intentions *after that time* [i.e., after a point in time on June 6, 2016], and if every maximally preferable option that is directly securable for Sally *on June 6, 2016* involves refraining from killing her firstborn child, then she refrains from killing her firstborn child). And, given the description of the case, come what may, Sally will satisfy the antecedent of this conditional. Thus, the only way Sally can satisfy this conditional is by refraining from killing her firstborn child on June 6, 2016. MWSS therefore implies that, on June 6, 2011, Sally is under an obligation to refrain from killing her firstborn child five years later. (Ross 2012, 92; italics added)

The first thing to notice about *MWSS* is that the times "for all t' from t forward" refer to the times when an agent has an obligation and that the obligatory act, *Φ*-ing, is implicitly also indexed to a time. It's important to be clear about this distinction, then, in order to accommodate the fact that we may have the same kind of obligation at *different* times to

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

do something at some future time. For example, possibilism, actualism, and securitism agree that in *WI1* Alice has both a t_2 obligation and a t_3 obligation to <K> at t_4.

Now, Ross's claim that *MWSS* avoids the nonratifiability problem relied on a claim about Sally's 2016 obligation (notice the parts that we italicized), notwithstanding the fact that Ross writes "on June 6, 2011, Sally ought to be such that, on June 6, 2016" which is not captured in the definition of *MWSS*, nor does it make any difference to what is being claimed since it is true at *all* times that Sally has a 2016 obligation to refrain from killing her firstborn child in 2016.[30]

When we turn to ask what, according to *MWSS*, Alice's 2011 obligation is, *MWSS* seems to imply that, on June 6, 2011, Sally is obligated to cut off her child's head on June 6, 2016. This is because, *at all times*, Sally ought to be such that on June 6, 2011, she satisfies the following conditional: (if whether Sally refrains from killing her firstborn child on June 6, 2016, does not depend on her intentions after June 6, 2011, and if every maximally preferable option that is directly securable for Sally on June 6, 2011, involves refraining from killing her firstborn child, then she refrains from killing her firstborn child). On June 6, 2011, Sally satisfies this conditional by failing to satisfy the second conjunct of the antecedent: no maximally preferable option that is directly securable for Sally on June 6, 2011, involves refraining from killing her firstborn child. Since Sally need (p. 155) not satisfy the conditional by satisfying the consequent, *MWSS* does not seem to imply that, on June 6, 2011, Sally is obligated to refrain from killing her firstborn child on June 6, 2016. Instead, *MWSS* seems to imply that, on June 6, 2011, Sally is obligated to kill her firstborn child on June 6, 2016. This is because Sally ought to be such that on June 6, 2011, she satisfies the following conditional: (if whether Sally kills her firstborn child on June 6, 2016, does not depend on her intentions after June 6, 2011, and if every maximally preferable option that is directly securable for Sally on June 6, 2011, involves killing her firstborn child, then she kills her firstborn child). Since the antecedent of this conditional is true, the only way for Sally to satisfy this conditional is by satisfying the consequent. Hence, *MWSS* seems to imply that, on June 6, 2011, Sally *is* obligated to kill her child on June 6, 2016. So, like securitism, *MWSS* similarly succumbs to the nonratifiability problem. In order to cement this conclusion, let's consider what *MWSS* says about *WI4*.

MWSS says that Alice ought to be such that, at t_2, she satisfies the following conditional: (if whether Alice <L-s> at t_4 does not depend on her intentions after t_2, and if every maximally preferable option that is directly securable for Alice at t_2 involves <L-ing> at t_4, then she <L-s> at t_4). Since Alice satisfies the antecedent of this conditional, the only way to satisfy the conditional is by satisfying its consequent. So, *MWSS* implies that Alice has a t_2 obligation to <L> at t_4. However, *MWSS* also implies that Alice has a t_3 obligation to <K> rather than <L> at t_4 because Alice satisfies the antecedent of the following conditional, (if whether Alice <K-s> at t_4 does not depend on her intentions after t_3, and if every maximally preferable option that is directly securable for Alice at t_3 involves <K-ing> at t_4, then she <K-s> at t_4), and so the only way to satisfy the conditional is by satis-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

fying the consequent. So, it seems that, like actualism and other forms of securitism, Ross's version of securitism is, in fact, subject to the nonratifiability problem.

The nonratifiability problem is, we believe, underexplored in the literature. This may be because actualists and securitists regard the problem as a merely quirky implication of their view, a small bullet to bite at best. However, we believe the nonratifiability problem points to a much deeper issue. We will argue that any view subject to the nonratifiability problem either violates "ought" implies "can" (OIC) or is committed to an implausible position on the relationship between dependent and nondependent obligations. As such, any view subject to the nonratifiability problem should be rejected.

Consider a case, *WI5*, that is just like *WI4*, except that we add the following stipulations: Alice is a fully rational agent from t_1-t_2,[31] and she knows from t_1-t_2 that, no matter what happens up until t_3, she won't have a t_3 obligation to <L> at t_4. Given these stipulations, it follows that Alice cannot, at t_1, <M> at t_2 if we embrace the following principle:

> **The Rationality-Ability Principle:** A fully rational agent S cannot, at t, decide at t^* ($t < t^*$) to perform an action x at t^{**} ($t^* < t^{**}$) if S knows from t - t^* that, no matter (p. 156) what happens up until t^{**}, S will have, after t^* and prior to t^{**}, decisive reason to refrain from x-ing at t^{**}.

Given the stipulations of *WI5* and the truth of the Rationality-Ability Principle, it follows in *WI5* that, at t_1, Alice cannot <M> at t_2 because part of the content of <M> includes playing video games at t_4, and Alice knows that, no matter what happens up until t_3, Alice lacks a t_3 obligation to play (and decide to play) video games at t_4. Despite Alice's inability at t_1 to <M> at t_2, it seems at first glance that securitism is committed to the position that Alice has a t_1 obligation to <M> at t_2 which would violate OIC. But suppose the securitist says in response that, since OIC is obviously true, it must not be the case that Alice has such a t_1 obligation after all. Then what, we may ask, is Alice obligated at t_1 to do at t_2 according to securitism?

Can securitists claim that Alice has a t_1 obligation to <C> at t_2? If Alice has a dependent t_1 obligation to <C> at t_2, this must be in virtue of Alice's nondependent t_1 obligation to <C, F, L> from t_2-t_4, even though the content of <C> includes doing research for a paper and *not* playing video games. However, despite doing the right thing at both t_2 and t_3, Alice has a t_3 obligation to <K> rather than <L> at t_4, and, unlike <L>, the content of <K> is shared by the content of <C>. The shared content between <C> and <K> is exactly what we would expect *if*, contrary to what securitists claim at this point in the dialectic, Alice has a nondependent t_1 obligation to <C, F, K>.

Goldman's version of securitism (1978, 194–195) which, as discussed in the previous section is, strictly speaking, not a version of maximalism, appears to rule out this response since, in *WI5*, although the act-set <C, F, L> is securable for Alice at t_1 in virtue of the fact that the performance of <C> would result in <C, F, L>, <C> would *not* be causally efficacious if performed in the sense that part of the content of <C> (doing research for a paper at t_4) would *not* be actualized if Alice were to <C> at t_2. Moreover, although

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Portmore's (2011, 193–194; 2019, chap. 5, fn. 10) version of securitism is more relaxed insofar as an act-set only needs to be securable in virtue of an immediately performable decision whose content is quite general,[32] it is unclear whether Portmore would want to allow for such a mismatch between the content of <C> and the content of <L>.

Perhaps securitists would say that in *WI5* Alice has a nondependent t_1 obligation to <C, F> from t_2-t_3 as opposed to <C, F, K> from t_2-t_4, thereby avoiding the mismatch between the content of <C> and the content of Alice's nondependent t_1 obligation. However, this move appears untenable because, according to securitism, the act-set <C, F> is obligatory for Alice at t_1 at least partly because of its value, and its value is at least partly determined by the fact that if Alice were to <C, F> from t_2-t_3 then she would <L> at t_4. To illustrate this point further, suppose that the value of <L> is –100 as opposed to +5. In that case, according to securitism, Alice would have a nondependent t_1 obligation to <B, D, I> from t_2-t_4 because this would be the best maximal act-set that is securable for Alice at t_1.

(p. 157) A final way in which securitists might respond to this alleged dilemma is by simply denying that Alice has a t_1 obligation to do anything at t_2, and that the lack of an obligation is explained at least in part by Alice's special knowledge of the fact that, no matter what she does up until t_3, she will lack a t_3 obligation to <L> at t_4. We think that a solution to the A/P debate should be able to handle the stipulations in *WI5* while affirming that Alice has a t_1 obligation to do something at t_2. For instance, possibilists have no difficulties maintaining that in *WI5* Alice has a t_1 obligation to <B> at t_2, and no knowledge of the kind that Alice has in *WI5* threatens this verdict. For example, if we further stipulate that in *WI5* Alice knows that, no matter what she does up until t_3, she lacks a t_3 obligation to <I> at t_4, this is no obstacle to the possibilist's contention that Alice has a t_1 obligation to <B> at t_2.

We don't declare possibilism victorious, however, since it is subject to the worst outcome objection. Instead, we propose that an alternative view, hybridism, avoids the nonratifiability problem as well as the other problems that plague actualism and possibilism.

8. Hybridism

The preceding discussion suggests that important *desiderata* for the correct view in the A/P debate include (i) accommodating the intuitively correct verdicts rendered by both actualism and possibilism, while avoiding the (ii) not demanding enough (iii) bad behavior (iv) and worst outcome objections and, perhaps most importantly, (v) the nonratifiability problem. For reasons already given, a view that can accommodate (i)–(v) should be particularly appealing to consequentialists. The good news for consequentialists is that there are such views, which we refer to collectively as hybridism. In this final section, we'll first provide a formulation of a particular version of hybridism and then explain how hybridism captures the aforementioned *desiderata*. These considerations provide good reason for consequentialists to accept hybridism over its competitors.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Hybrid views are unique, in part, because they posit two distinct moral "oughts," one actualist in nature and one possibilist in nature. These oughts are meant to jointly track the insights of both actualism and possibilism, yet be immune from the five aforementioned objections. Given space limitations, we cannot provide a complete defense of any particular hybrid view here, so our goal is merely to make a prima facie case for hybridism by explaining how it satisfies the aforementioned *desiderata*. We will focus on one version of hybridism known as *single obligation hybridism* (*SOH*). In its simplest form, *SOH* posits a possibilist moral *obligation* that picks out the criterion of right, which allows *SOH* to (a) accommodate the intuitively correct possibilist verdicts and to (b) avoid the not demanding enough objection (c) the bad behavior objection and (d) the nonratifiability problem. In order to (a) accommodate the intuitively correct actualist verdicts and (b) avoid the worst outcome objection, *SOH* also posits an actualist moral *ought* that functions as a sort of decision procedure. In certain cases, it prescribes (p. 158) performing a wrong act at one time in order to avoid performing an even worse act at a different time. This moral ought is an action-guiding practical ought, not a moral obligation. *SOH* may be formulated more precisely as follows (Timmerman and Cohen 2016, 682–683).

> ***Single Obligation Hybridism:***
>
> **Possibilist Moral Obligation:** At *t* an agent *S* has a possibilist moral obligation to *φ* at *t'* iff *φ*-ing at *t'* is part of the best series of acts that *S* can perform from *t* to the last moment that *S* can possibly perform an act.
>
> **Actualist Practical Ought:** At *t* an agent *S* has most practical reason to φ at *t'* iff φ-ing at *t'* is under *S*'s control at *t* and φ-ing at *t'* is either (i) identical to the maximally specific possibilist obligation that *S* has at *t*, (ii) a rationally permissible supererogatory act, or (iii) is the least rationally impermissible, all things considered, act-set presently under S's control at *t*. There is an act-set that satisfies (iii) iff no act-set presently under S's control at t satisfies conditions (i) or (ii).
>
> At t an agent S has a dependent possibilist obligation to perform an act-set *Ψ* iff (and because) at t S has a nondependent possibilist moral obligation to perform an act-set *Ψ**, such that *Ψ* is a proper subset of *Ψ**. Additionally, at t, S dependently has most practical reason to perform act-set *Ψ* iff (and because) at t S nondependently has most practical reason to perform an act-set *Ψ**, such that *Ψ* is a proper subset of *Ψ**.

The technical details of *SOH* are not centrally important for the purposes of this chapter. The most important elements of the view may be understood by considering the nontechnical description of the view that preceded the formal definition and by considering *SOH*'s applications in particular cases. Consider *WI1* once more. *SOH* entails that Alice has a possibilist moral obligation to <B, D, H> because this is the best maximal act-set that Alice can perform over time. This is the intuitively correct verdict. This feature allows *SOH* to avoid both the not demanding enough and bad behavior objections for obvious reasons. It also allows *SOH* to avoid the nonratifiability problem because Alice will always be obligated to perform each basic act that is part of the best set of acts she can

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

perform at any given time. This possibilist feature of *SOH* makes it logically impossible for Alice to, at one time, have an obligation to perform an act at some future time and then, after only fulfilling her obligations, cease to have the future obligation in question. There is thus no reason to worry about *SOH* succumbing to the nonratifiability problem.

At the same time, *SOH* is immune from the worst outcome objection because the actualist ought, not the possibilist obligation, is action-guiding. The actualist ought prescribes performing the act that would result in the best outcome *from among the set of acts presently under the agent's control*. This practical ought, then, serves the purpose of minimizing wrongdoing in light of one's present circumstances. So, in *WI1*, *SOH* entails that Alice practically ought to <C>, which would result in her performing <C, F, K>. This captures the intuitively correct actualist verdict. *SOH* has this implication because, (p. 159) of the act-sets at t_2 under Alice's control at t_1, performing <C> would result in the best outcome. So, *SOH* entails that Alice ought to perform a wrong act now (i.e., <C>) in order to prevent herself from performing an even worse act later (i.e., <I>). There is much more to be written in favor of (and against) hybridism. Indeed, there is much more to be written about each of these views. The considerations we raise in this section, however, should provide a least a presumptive case in favor of hybridism.

9. Conclusion

This chapter served a few related goals. We first traced the origins of the A/P debate to a debate in the deontic logic literature about the coherence of utilitarianism. We then discussed the relationship between time, an agent's abilities, and her moral obligations before introducing a precise version of the wedding invitation case that would guide us through the chapter. Using that case, we reviewed the four primary views in the A/P debate, explaining their benefits and drawbacks from a consequentialist perspective. In doing so, we attempted to further the A/P debate by arguing that all forms of securitism (including Ross's *MWSS*) are subject to the nonratifiability problem. Moreover, we argued that the nonratifiability problem is more serious than it may initially seem, as it either violates OIC or is committed to an implausible position on the relationship between dependent and nondependent obligations. We ended by making a positive case for hybridism which, we argued, avoids the nonratifiability problem in addition to each of the other problems discussed in the chapter. These considerations suggest that hybridist forms of consequentialism will be the most plausible forms of consequentialism.[33]

References

Almeida, M. 1992. "The Paradoxes of Feldman's Neo-Utilitarianism." *Australasian Journal of Philosophy* 70, no. 4: 455–468.

Baker, D. 2012. "Knowing Yourself–and Giving Up On Your Own Agency in The Process." *Australasian Journal of Philosophy* 90, no. 4, 641–656.

Bergström, L. 1968. "Utilitarianism and Deontic Logic." *Analysis* 29, no. 2: 43–44.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Brown, C. 2018. “Maximalism and the Structure of Acts.” *Noûs* 52, no. 4: 752–771.

Bykvist, K. 2002. “Alternative Actions and the Spirit of Consequentialism.” *Philosophical Studies* 107, no. 1: 45–68.

Cariani, F. 2016. “Consequence and Contrast in Deontic Semantics.” *Journal of Philosophy* 113, no. 8: 396–416.

Carlson, E. 1995. *Consequentialism Reconsidered*. Dordrecht, the Netherlands: Kluwer.

Castañeda, H. 1968. “A Problem for Utilitarianism.” *Analysis* 28, no. 4: 141–142.

(p. 160) Cohen, Y., and Timmerman, T. 2016. “Actualism Has Control Issues.” *Journal of Ethics and Social Philosophy* 10, no. 3: 1–19.

Dorsey, D. 2009. “Headaches, Lives and Value.” *Utilitas* 21, no. 1: 36–58.

Dreier, J. 2011. “In Defense of Consequentializing.” In *Oxford Studies in Normative Ethics, volume 1*, edited by M. Timmons, 97–119. New York: Oxford University Press.

Feldman, F. 1986. *Doing the Best We Can: An Essay in Informal Deontic Logic*. Boston: D. Reidel.

Goldman, H. S. 1976. “Dated Rightness and Moral Imperfection.” *The Philosophical Review* 85, no. 4: 449–487. [Now Holly M. Smith]

Goldman, H. S. 1978. “Doing the Best One Can.” In *Value and Morals: Essays in Honor of William Frankena, Charles Stevenson, and Richard Brandt*, edited by A. I. Goldman and J. Kim, 185–214. Dordrecht, the Netherlands: D. Reidel. [Now Holly M. Smith]

Greenspan, P. S. 1978. “Oughts and Determinism: A Response to Goldman.” *The Philosophical Review* 87, no. 1: 77–83.

Gustafsson, J. E. 2014. “Combinative Consequentialism and the Problem of Act Versions.” *Philosophical Studies* 167, no. 3: 585–596.

Jackson, F. 1985. “On the Semantics and Logic of Obligation.” *Mind* 94, no. 374: 177–195.

Jackson, F., and Pargetter, R. 1986. “Ought, Options, and Actualism.” *The Philosophical Review* 95, no. 2: 233–255.

Kiesewetter, B. 2018. “Contrary-to-Duty Scenarios, Deontic Dilemmas, and Transmission Principles.” *Ethics* 129, no. 1: 98–115.

Lewis, D. 1973. *Counterfactuals*. Oxford: Blackwell. [Reprinted with revisions, 1986].

Norcross, A. 1998. “Great Harms from Small Benefits Grow: How Death Can Be Outweighed by Headaches.” *Analysis* 58, no. 2: 152–158.

Pollock, J. L. 1976. *Subjunctive Reasoning*. Dordrecht, the Netherlands: D. Reidel.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Portmore, D. W. 2009. “Consequentializing.” *Philosophy Compass* 4, no. 2: 329–347.

Portmore, D. W. 2011. *Commonsense Consequentialism*. New York: Oxford University Press.

Portmore, D. W. 2017. “Maximalism versus Omnism about Permissibility.” *Pacific Philosophical Quarterly* 98(S1): 427–452.

Portmore, D. W. 2019. *Opting for the Best: Oughts and Options*. New York: Oxford University Press.

Prawitz, D. 1968. “A Discussion Note on Utilitarianism.” *Theoria* 34, no. 1: 76–84.

Prawitz, D. 1970. “The Alternatives to Action.” *Theoria* 36, no. 2: 116–126.

Ross, J. 2012. “Actualism, Possibilism, and Beyond.” In *Oxford Studies in Normative Ethics, volume 2*, edited by M. Timmons, 74–96. New York: Oxford University Press.

Sobel, J. H. 1976. “Utilitarianism and Past and Future Mistakes.” *Noûs* 10, no. 2: 195–219.

Stalnaker, R. 1968. “A Theory of Conditionals.” In *Studies in Logical Theory* (American philosophical quarterly monograph series: volume 2), edited by N. Rescher, 98–112. Oxford: Blackwell.

Timmerman, T. 2015. “Does Scrupulous Securitism Stand-up to Scrutiny? Two Problems for Moral Securitism and How We Might Fix Them.” *Philosophical Studies* 172, no. 6: 1509–1528.

Timmerman, T. 2019. “Effective Altruism’s Underspecification Problem.” In *Effective Altruism: Philosophical Issues*, edited by H. Greaves and T. Pummer, 166–183. New York: Oxford University Press.

Timmerman, T., and Cohen, Y. 2016. “Moral Obligations: Actualist, Possibilist, or Hybridist? *Australasian Journal of Philosophy 94*(2), 672–686.

(p. 161) Timmerman, T., and Cohen, Y. 2019. “Actualism and Possibilism in Ethics.” In *The Stanford Encyclopedia of Philosophy* (Summer 2019 edition), edited by E. N. Zalta. **https://plato.stanford.edu/archives/sum2019/entries/actualism-possibilism-ethics/**.

Timmerman, T., and Swenson, P. 2019. “How to Be an Actualist and Blame People.” In *Oxford Studies in Agency and Responsibility, volume 6*, edited by D. Shoemaker, 216–240. New York: Oxford University Press.

Vessel, J. 2009. “Defending a Possibilist Insight in Consequentialist Thought.” *Philosophical Studies* 142, no. 4: 183–195.

Vessel, J. 2016. “Against Securitism, the New Breed of Actualism in Consequentialist Thought.” *Utilitas* 28, no. 2, 164–178.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Wedgwood, R. 2009. “Against Actualism.” PEA Soup, posted on September 11, 2009. **http://peasoup.typepad.com/peasoup/2009/09/against-actualism.html**

Westphal, F. 1972. “Utilitarianism and “Conjunctive Acts’: A Reply to Professor Castañeda.” *Analysis* 32, no. 3: 82–85.

Woodard, C. 2009. “What’s Wrong with Possibilism.” *Analysis* 69, no. 2: 219–226.

Zellner, H. (1972). “Utilitarianism and Derived Obligation.” *Analysis*, 32 no. 4, 124–125.

Zimmerman, M. J. 1986. “Subsidiary Obligation.” *Philosophical Studies* 50, no. 1: 65–75.

Zimmerman, M. J. 1990. “Where Did I Go Wrong?” *Philosophical Studies* 59, no. 1: 55–77.

Zimmerman, M. J. 1996. *The Concept of Moral Obligation*. Cambridge: Cambridge University Press.

Zimmerman, M. J. 2006. “The Relevant Risks to Wrongdoing.” In *The Good, the Right, Life and Death: Essays in Honor of Fred Feldman*, edited by K. McDaniel, J. R. Raibley, R. Feldman, and M. J. Zimmerman, 151–172. New York: Ashgate.

Zimmerman, M. J. 2017. “Prospective Possibilism.” *Journal of Ethics* 21, no. 2: 117–150.

Notes:

([1]) Cf. Dreier (2011, 97) and Portmore (2009, 330; 2011, 34–38).

([2]) We will suppose that an agent S performs an act-set (singleton or plural) x at time t iff x is a possible state of affairs involving an action (or a number of actions) that belong(s) to S and x is actualized at t.

([3]) When such cases are discussed in the literature, the focus of the disagreement is about whether there is a lexical priority between values. See Norcross (1998) and Dorsey (2009).

([4]) These issues also feature centrally in certain debates about deontic logic and the related maximalism/omnism debate, which we’ll briefly discuss in section 2.

([5]) This case is drawn from Zimmerman (2006, 153). Since this chapter is focused on the actualism/possibilism debate as it applies to consequentialism, the ranking of outcomes should be understood to be determined by intrinsic value, as opposed to deontic value.

([6]) The A/P debate has typically focused on CFs in which “C” includes S’s performing an action at t_1. It also seems to conceive of freedom in terms of having an ability to do otherwise. This is because everyone in the debate implicitly assumes that an agent can do something, such that if they were to do it, then some true CF would be false instead (Cohen and Timmerman 2016, 4).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([7]) The circumstances “C” must refer to (temporally intrinsic) maximally specified circumstances at some time in order to accommodate the fact that adding information into the antecedent of a counterfactual can alter its truth value (Jackson 1985, 178, 186; Lewis 1973; Stalnaker 1968).

([8]) Lars Bergström (1968, 43) responded to Castañeda’s argument against utilitarianism by pointing out that the contradiction arises only if it is assumed that <A> and <B> are *alternatives* in the relevant sense, but <A> and <B> are not, in fact, alternatives since the agent can consistently perform both of them.

([9]) Notably, Prawitz (1968; 1970) and Westphal (1972) each argued that an act is permissible if and only if it is part of an act-set that, if performed, would bring about the greatest net good of any of the act-sets available to the agent. Prawitz and Westphal were essentially giving what may be considered the earliest defenses of possibilism, though they did not refer to this view as such.

([10]) An action that is uniquely optimific at a time is one that is optimific relative to all other performable actions at that time.

([11]) We agree with Portmore and Brown that, more exactly, the object of a nondependent obligation is a maximally normatively specific option (or act-set). This is an act-set that is entailed only by normatively equivalent options/act-sets (Brown 2018, 13; Portmore 2017, 428). All of the arguments put forward in this paper regarding the status of maximal act-sets similarly apply, mutatis mutandis, to the status of maximally normatively specific options/act-sets.

([12]) Cf. Bykvist (2002, 47), Carlson (1995, 77), Goldman (1976, 453), and Portmore (2011, 166).

([13]) The ability to perform a (nonsingleton) act-set over time may be understood in terms of having an ability to perform the first act in the act-set, and then, once the first act is performed, she will have the ability to perform the second act in the act-set, and once the second act is performed, she will have the ability to perform the third act, etc., culminating in the agent’s ability to perform the final act in that act-set. See, e.g., Goldman (1978, 193) and Portmore (2011, 165–166, 170).

([14]) This notably includes consequentialist effective altruists. See Timmerman (2019).

([15]) Feldman (1986, 41–44); Goldman (1978, 80); Kiesewetter (2018); Vessel (2009); Zimmerman (1990, 58–60); Zimmerman (2006, 154–155).

([16]) There is more to be said in favor of possibilism that is beyond the scope of this paper. See Zimmerman (1996, fn. 72 and fn. 122) and (2017, chap. 3) for a nice review of some of possibilism’s additional, lesser appreciated, virtues.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([17]) Almeida (1992, 461–462); Feldman (1986, 52–57); Goldman (1976, 469–470); Gustafsson (2014, 593); Ross (2012, 81–82); Sobel (1976, 202–203); Timmerman and Cohen (2016, 674); Woodard (2009, 219–221).

([18]) See Greenspan (1978, 81) and Zimmerman (1986, 70; 2017, 126–128).

([19]) This formulation of actualism avoids what is often referred to as the lumping problem (Cariani 2016; Wedgwood 2009). A different formulation of actualism may be found in Jackson and Pargetter (1986). For a detailed overview of their version, and how it is different from alternative versions, see Timmerman and Cohen (2019).

([20]) Note that all maximal act-sets are fully specified act-sets, but not all fully specified act-sets are maximal act-sets. For example, suppose that an agent has, at t_1, an ability to perform a fully specified act-set <x> at t_2. Hence, there is no act-set that is not identical to <x> that, at t_1, the agent can perform *at* t_2 that contains <x>. However, <x> itself is not maximal if it is contained in the following maximal act-set that the agent can, at t_1, perform *over time* from t_2-t_3: <x, y>.

([21]) This counterfactual is true since, as a matter of stipulation, both its antecedent and its consequent are true (Lewis 1973, 132; Pollock 1976, 42–43; Stalnaker 1968).

([22]) See Goldman (1978, 202) and Portmore (2011, 179). Gustafsson's (2014) solution to Castañeda's (1968) formal critique of consequentialism appeals to jointly exhaustive and mutually exclusive act-sets. Gustafsson considers his position to be distinct from maximalism, but Brown (2018, 66–67) persuasively argues that Gustafsson's position is at least consistent with maximalism.

([23]) Sobel (1976) refers to a maximal act-set as a "life."

([24]) Cf. Bykvist (2002, 50–51).

([25]) However, see Timmerman and Swenson (2019) for an argument that possibilism is subject to an analogue problem.

([26]) We take it that some actions are more difficult to perform than others. Fill out this account whichever way you like and suppose that each securable act is easy to perform in this sense.

([27]) Baker (2012, 642–643); Jackson and Pargetter (1986, 240); Portmore (2011, 207); Timmerman (2015, 1512–1513); Zimmerman (2006, 156).

([28]) This is not strictly entailed by actualism, but it is entailed by actualism coupled with widely accepted axioms about moral blameworthiness. For more on the relationship between actualism, possibilism, and blame, see Timmerman and Swenson (2019).

([29]) Ross (2012); Timmerman and Cohen (2016); Wedgwood (2009); Zimmerman (2017), 121.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([30]) A bit more precisely, Sally has an obligation at t (which is in 2016) to decide at some time t^* after t (which is also in 2016) to refrain from killing her firstborn child (also in 2016).

([31]) One might object that Alice must not be a fully rational agent from t_1-t_2 in *WI5* because she performs <C> at t_2, and, for some reason, this is not what a fully rational agent would do in these circumstances. Even if this is right, we can add the further stipulation that *WI5* remains silent on what Alice in fact does at t_2.

([32]) Cf. Brown (2018, 764–766) and Gustafsson (2014, 587–588).

([33]) We are grateful to Doug Portmore for very helpful feedback on this chapter. This chapter is the product of full and equal collaboration between its authors.

Yishai Cohen

Yishai Cohen is Assistant Professor of Philosophy at the University of Southern Maine. His research focuses on agency, ethics, metaphysics, and the philosophy of religion. He is particularly interested in the relationship between libertarian free will and a variety of issues in ethics, including 'Ought' Implies 'Can', the Principle of Alternate Possibilities, and the actualism/possibilism debate.

Travis Timmerman

Travis Timmerman is Assistant Professor of Philosophy at Seton Hall University. He specializes in normative ethics, applied ethics, and the philosophy of death. In normative ethics, he primarily works on the actualism/possibilism debate, having recently coauthored the Stanford Encyclopedia of Philosophy (2019) entry on the topic as well as "How To Be an Actualist and Blame People" in Oxford Studies in Agency and Responsibility (2019). In the death literature, he focuses on axiological questions about death's badness. Recent publications include "A Dilemma for Epicureanism" in Philosophical Studies (2019) and "Avoiding the Asymmetry Problem" in Ratio (2018). In applied ethics, he works on issues related to global poverty and questions about the ethics of historical monuments. Publications in applied ethics include "Sometimes There Is Nothing Wrong with Letting a Child Drown" in Analysis (2015) and "A Case for Removing Confederate Monuments" in Oxford University Press's Ethics Left and Right (2020).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Relativized Rankings

Matthew Hammerton

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.6

Abstract and Keywords

In traditional consequentialism the good is position-neutral. A single evaluative ranking of states of affairs is correct for everyone, everywhere regardless of their positions. Recently, position-relative forms of consequentialism have been developed. These allow for the correct rankings of states to depend on connections that hold between the state being evaluated and the position of the evaluator. For example, perhaps being an agent who acts in a certain state requires me to rank that state differently from someone else who lacks this connection. In this chapter several different kinds of position-relative rankings related to agents, times, physical locations, and possible worlds are explored. Arguments for and against adopting a position-relative axiology are examined, and it is suggested that position-relative consequentialism is a promising moral theory that has been underestimated.

Keywords: position-relative, agent-relative, agent-neutral, time-relative, world-relative, location-relative, deontic constraints, special duties, consequentializing

ALL consequentialist theories contain an axiology—a theory of the good that ranks various outcomes as better or worse. Although consequentialism has historically been associated with certain axiologies, there are in principle no restrictions on the axiological structures a consequentialist theory can contain. This chapter is concerned with one important distinction in axiological structure: the distinction between position-neutral and position-relative axiologies. When a state of affairs is evaluated, there are various connections that may hold between the position of the evaluator and the state evaluated. For example, the evaluator may be an agent who acts in the state she is evaluating. Or she may have the same temporal location as the state she is evaluating. In a *position-neutral axiology* these kinds of connections are irrelevant to the correct ranking of states. In a *position-relative axiology* these kinds of connections are sometimes relevant to the correct ranking. In other words, the rankings are indexed to certain positional facts such that a certain ranking may be correct relative to one position but not relative to another. Different kinds of position-relative rankings are available depending on what positional facts are relevant. *Agent-relative* rankings make certain connections between the numerical identity of the

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

evaluator and the state evaluated relevant. *Time-relative* rankings make certain connections between the temporal location of the evaluation and the temporal location of the evaluated state relevant. Further positional facts related to spatial location, possible world location, and patient identity may also result in relativized rankings and are discussed in this chapter.

It is notable that position-relative axiologies have been neglected for much of the history of consequentialism. The classical utilitarians did not recognize the possibility of position-relative value and assumed by default that all value is position-neutral. Subsequent generations of consequentialists followed them in making this assumption and it was not until Amartya Sen published "Rights and Agency" in 1982 that a position-relative form of consequentialism was first proposed.[1] By contrast, several other significant (p. 47) axiological assumptions made by the classical utilitarians were challenged much earlier in the history of consequentialism. For example, the classical utilitarians adopted an axiology that is monadic (there is only one intrinsic good), welfarist (all goods are good for someone), and atomistic (the good of a whole is a sum of the good of its parts). Yet by 1903 a form of consequentialism that abandons each of these axiological assumptions had already been put forward by G. E. Moore in his *Principia Ethica*. The delayed recognition of position-relative consequentialism may be explained by the fact that, more so than any other axiological structure, position-relative rankings appear to lack a precedent in everyday moral talk and thought. Some have used this to argue that position-relative values are implausible, while others have disputed the claim that they are unprecedented. These issues are discussed in section 3.1.

There are many reasons why consequentialists are interested in position-relative rankings. However, one reason stands out above all others. It has long been held that consequentialism, despite its axiological openness, is unable to capture commonsense moral verdicts involving deontic constraints and special duties. On these grounds many have rejected consequentialism despite finding its theoretical framework attractive. If position-relative consequentialism is viable, then it appears to be a game changer in this debate. For position-relative consequentialism can produce deontic constraints and special duties. Thus, it opens the way for an attractive form of consequentialism that preserves key aspects of the consequentialist framework yet captures the intuitive verdicts of commonsense morality.

One final thing to note. This article will follow the dominant contemporary practice of extending the term "consequentialism" to all theories that tie the deontic status of acts (or rules or motives) to the promotion of outcomes ordered by some evaluative ranking. Some have insisted on restricting the term "consequentialism" to theories with neutral evaluative rankings because of the strong historical associations between consequentialism and agent-neutrality.[2] However, we need not get caught up in a debate about how to use the word "consequentialism." What ultimately matters is not the terms we use but rather that we clearly see the logical space of theoretical options. Those who prefer the restricted use of "consequentialism" should still admit that there is an important family of theories in normative ethics that borrow key structural features from traditional conse-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

quentialism yet employ a position-relative axiology. This family deserves careful examination in a volume like this.

1. Agent-Relativity

When Sen (1982) first proposed a position-relative form of consequentialism, he was proposing an agent-relative form. Although other categories of relative value have since (p. 48) been explored, the agent-relative variety is by far the most significant and most widely discussed. Therefore, this section will focus on agent-relative value, with other varieties of relative value being explored in the next section.

1.1. Agent-Relative Rules and Theories

Before examining agent-relative value, we should first consider what it is for a moral rule, reason, or theory to be agent-relative. I will use Derek Parfit's (1984) intuitive account of the distinction, although several formal accounts are also available.[3] Parfit takes all moral rules to give agents certain substantive aims. For example, a rule that prohibits lying gives each agent the aim that she does not lie. Parfit defines an agent-neutral rule as one that gives all agents the same aim and an agent-relative rule as one that gives different aims to different agents. So, for example, consider the following rules:

(1) Each agent must not tell lies.
(2) Each agent must ensure, to the best of her ability, that her family is honest.
(3) Each agent must ensure, to the best of her ability, that everyone is honest.
(4) Each agent must minimize general violations of (1).

Rules (1) and (2) are generally regarded as agent-relative rules, whereas (3) and (4) are generally regarded as agent-neutral. Parfit's account explains this as follows. Rules (1) and (2) are agent-relative because they give different aims to different agents. For example, (1) gives Akira the aim that *Akira does not tell lies* and Mia the different aim that *Mia does not tell lies*. Likewise, (2) gives me the aim that *my* relatives are honest and you the different aim that *your* relatives are honest. By contrast, (3) and (4) give the same aims to all agents. For example, (3) gives you and me (and everyone else) the common aim that everyone is honest. Likewise, (4) gives everyone the common aim that no one violates (1), and that if there are any violations of (1), then they are minimal.[4]

Parfit extends his account to moral theories. He says that an *agent-neutral theory* gives all agents the same aims, whereas an *agent-relative theory* sometimes gives different aims (p. 49) to different agents. It follows that any theory containing at least one agent-relative rule will itself be an agent-relative theory, whereas only theories with exclusively agent-neutral rules are agent-neutral.

Understanding the agent-relative/neutral distinction helps us to clarify two important features of commonsense morality: *deontic constraints* and *special duties*. A deontic constraint prohibits agents performing acts of a certain type even if doing so is the only way

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

to prevent more acts of that type being performed by others. A special duty requires agents to give priority to people that they have some kind of special relationship with (e.g., parent and child, siblings, friends, compatriots, etc.), even if not giving this priority would result in more prioritization of special relationships by others. Let's consider examples of each. First, consider the following case:

> MURDER: Belle is in a murderous rage and is about to kill two innocent people. The only way to stop her is for you to kill a random innocent person (who is not one of Belle's potential targets).

Many people judge that it would be wrong for you to kill the innocent person in this case even though it will result in fewer innocent people being killed overall. In making this judgment, they are endorsing a deontic constraint on killing innocent people.

Second, consider the following case:

> ANTIVENOM: Kimbo's daughter and two other children he has no ties to have all been bitten by a brown snake. Kimbo has one dosage of antivenom available. To save his daughter's life, he will need to give her a full dosage. The other two children can each be saved with a half dosage. Kimbo could allow the parents of these children to administer a half dosage to each child, in which case their children live and his daughter dies. Or he could administer the full dosage to his daughter, in which case she lives and the other children die.

Many people judge in this case that Kimbo ought to administer the antivenom to his own child, even though offering it to the other parents would result in more children being saved by their parents. In making this judgment they are endorsing a special duty requiring parents to favor their own children.

Deontic constraints and special duties are both agent-relative rules as they give different aims to different agents. For example, a deontic constraint on killing gives me the aim that *I do not kill* and you the aim that *you do not kill*. Likewise, a special duty for parents to favor their children gives me the aim that *I favor my child* and you the aim that *you favor yours*. Some have attempted an agent-neutral interpretation of deontic constraints and special duties by suggesting that a common aim could account for them. For example, Dougherty (2013) and Setiya (2018) have argued that killing in order to prevent more killings is a different act from killing for some other end such as revenge. A rule requiring us to prioritize minimizing the first kind of killing over minimizing the second kind gives us all a common aim and yet appears to produce a deontic constraint on killing. However, in Hammerton (2017) I show that this strategy is inadequate. Suppose (p. 50) that I can either ensure that I do not vengefully kill someone or ensure that you do not vengefully kill someone. The deontic constraint on killing requires me to ensure first that I do not kill, prioritizing my own nonkilling over yours. Yet this cannot be captured by the common aim that *no one kills to prevent more killings*. Only an agent-relative rule giving each agent the aim that *she does not kill* can capture this deontic verdict.[5]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Deontic constraints and special duties are core features of our commonsense morality (as our intuitive reactions to MURDER and ANTIVENOM demonstrate). As such, they show that commonsense morality is necessarily agent-relative. This presents a challenge to all agent-neutral theories. Such theories are incompatible with core features of commonsense morality and need to be defended against the charge that this incompatibility makes them implausible. Thus, the distinction between agent-neutral and agent-relative moral theories is thought to mark an important distinction between theories that can capture certain core features of commonsense morality and those that cannot.

When it was assumed that a consequentialist axiology is agent-neutral, the aforementioned distinction was thought to divide consequentialism from its main rivals, which all appeared to be agent-relative theories. However, the possibility of agent-relative rankings opens the way for agent-relative versions of consequentialism that accommodate deontic constraints and special duties. Thus, the possibility of agent-relative rankings is especially relevant to the question of whether consequentialism can accommodate these core features of commonsense morality.

1.2. Agent-Relative Rankings

The key idea behind agent-relative rankings is that an ordering of possible outcomes as better or worse could be relativized to agents. This relativization allows for the correct ranking of states to depend on the agent who is evaluating those states such that, relative to one agent, state S_1 might be correctly ranked *above* state S_2, whereas, relative to another agent, S_1 might be correctly ranked *below* S_2.

If such relativized rankings are adopted, then the value that they are supposed to capture must itself be relative. Things are not just morally good or bad, they are morally good or bad relative to particular agents. It follows that moral goodness is not a property possessed by states of affairs but a two-place relation between a state of affairs and an agent.[6]

(p. 51) Admitting agent-relative value into the axiology of a consequentialist theory allows it to capture deontic constraints and special duties. To see this, let’s consider the deontic constraint on killing. According to this constraint, each agent must not kill an innocent person even if it is the only way to prevent more killings by others. We saw earlier that an agent-neutral theory cannot capture this constraint. Yet a consequentialist theory with an agent-relative axiology can. It does this by holding that, for each agent, that *she* kills an innocent person is *worse-relative-to her* than any number of innocent people being killed by others.[7] The result is that when each agent maximizes what is *good-relative-to her*, she refrains from killing innocent people even in cases like MURDER where it can prevent more killings by others. A similar move can capture special duties. The consequentialist can hold that, for each agent, that she neglects her child is *worse-relative-to her* than several other parents neglecting their children. Thus, maximizing what is *good-relative-to her* requires each parent to not neglect her child even when doing so can prevent more child neglect by others.[8]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Relativized Rankings

Focusing on deontic constraints and special duties suggests that agent-relative value is based on a relation of identity between the evaluator of a state and an agent acting in that state.[9] For example, if I am comparing my killing an innocent person with two strangers killing an innocent person, I rank the former state as worse because I am identical to the agent who kills in that state. However, further examples suggest that other kinds of connections between the evaluator and the evaluated state could justify agent-relative rankings. For example, compare a state in which my mother drowns to one in which an equally worthy stranger drowns. Now suppose that you have no personal connections to either my mother or the stranger. How ought we to rank these states? Given my personal connection to my mother, it appears that I ought to rank her drowning as worse than a stranger drowning. Given your lack of personal connection, it seems that you ought to rank these states as equally bad. Neither of us acts in the states we are assessing. However, my special relationship with a moral patient whose welfare is at stake in these states (and your lack of such a relationship) requires us to rank these states differently.[10]

To better understand agent-relative value, there are several further points that we need to clarify. First, we must not confuse what is *good-relative-to* an agent with what is (p. 52) *good for* an agent.[11] Both are categories of value that connect an agent with a state of affairs. However, the former is a kind of *moral* value, whereas the latter is a kind of *prudential* value. Although some moral philosophers hold substantive views that connect moral value and prudential value (e.g., derive one from the other), they nonetheless remain conceptually distinct. This is why, in a situation where several people are drowning and either Azra will be rescued or five others will be rescued, it is not incoherent to say that "It is better for Azra that she is rescued, yet, from a moral perspective, it would be better if the five others were rescued instead." An agent-relative consequentialist might want to make similar distinctions. For example, in a situation where a murderer will kill Azra and another innocent person unless Azra herself kills a random innocent person, the consequentialist might want to say that Azra killing the innocent person is *better for* her, yet, from a moral perspective, it is *worse-relative-to* her. Even a consequentialist who does not want to say this because she accepts a tight connection between prudential goods and moral goods should at least admit that it is a possible view that some might want to hold because *good for* and *good-relative-to* are distinct concepts.

Second, that something is *good-relative-to* an agent does not entail that its value is constituted by, or dependent on, the mental states of that agent. Some philosophers who endorse agent-relative value hold that it is mind-dependent and some hold that it is mind-independent. Just as is the case with agent-neutral value, both positions are possible. However, although agent-relative value does not entail mind dependence, there are certain mind-dependent accounts of value that entail that value is agent-relative. We will look at this in more detail later when we discuss an important example given by Michael Smith.

Third, on the standard understating of agent-relative value, each agent is supposed to recognize the correctness of each set of rankings relative to the agent it is indexed to, rather than treating her own relativized rankings as the universal yardstick by which

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

everyone is assessed. Thus, if you act in a way that maximizes that which is *good-relative-to* you but does not maximize that which is *good-relative-to* me, then I must recognize you as acting correctly even though, from my evaluative perspective, things would have gone better if you had acted differently. A consequence of this is that it is only assessments of value that are relative and not assessments of an agent's actions or attitudes. Although different agents may correctly rank the same state of affairs differently, they must accept that there is only one correct deontic status for each specific act based on what is *good-relative-to* the agent who performs the act.

Fourth, some have found it illuminating to explain agent-relative value in terms of the fittingness analysis of value.[12] Independently of debates about agent-relative value, some have found the following analysis of value appealing:

FIT: A state of affairs is valuable if, and only if, it is fitting to desire its realization.

(p. 53) If this analysis is accepted, then it provides an attractive account of why value comes in agent-relative and agent-neutral varieties. For it is plausible to hold that which states it is fitting to prefer at least sometimes depends on who you are and what relations hold between you and the people or things in the states you are assessing. For example, we might think that it is fitting for a parent to prefer that her child has a pleasant experience over another child having an equally pleasant experience, yet fitting for a neutral bystander to be indifferent about which child has the experience. If our evaluative ranking of states is derived from what it is fitting to prefer, and what it is fitting to prefer can vary from agent to agent, then the correct evaluative ranking can also vary from agent to agent.

Several other analyses of value that are structurally similar to FIT also appear to explain agent-relative value. For example, analyzing value in terms of which states we have most reason to prefer, or which states it is appropriate to love for their own sake, or which states an idealized version of ourselves would desire all make room for relative values.[13] By contrast, the Moorean account of value—often regarded as the main rival of the fittingness analysis—is not as accommodating of agent-relative value. According to the Moorean account, value is unanalyzable—it cannot be explained in terms of other, more fundamental, normative properties. If the fact that a state is valuable is just a brute normative fact, then it is difficult to see why that value would be relativized to agents. Nonetheless, a Moorean might offer independent reasons for postulating relative values, such as their theoretical utility. These issues will be discussed further in section 3.1.

Finally, earlier we made the simplifying assumption that if value is relativized to agents, then different evaluative rankings will be correct for different agents. However, this assumption is false (see Smith 2003, 2009), and this has important implications for how we characterize agent-relative value. To see that it is false, consider a version of the dispositional theory of value holding that to judge that a state S_1 is better than S_2 is to judge that you would prefer S_1 to S_2 if you had a set of preferences that were maximally informed, coherent, and unified. This analysis of value entails that claims about value are indexed to agents because they are analyzed in terms of what the person making the value judge-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ment would ideally prefer. Yet it does not entail that different rankings are correct relative to different agents. For it could be the case that the idealized preferences of all possible agents converge, and if they did coverage, then a single evaluative ranking would be correct for all agents. Thus, holding the dispositional theory of value, and endorsing convergence among all possible agents, results in a theory in which the "better-than relation" is relativized to agents and yet there is a single correct evaluative ranking for all agents.

This example shows us that indexing evaluations to agents is not sufficient for there to be different correct evaluative rankings for different agents. However, the former is at least necessary for the latter because, if an evaluative ranking is correct for one agent but not correct for another, then that ranking must be indexed to the first agent but not the second.

(p. 54) Given this result, we need to clarify what we are trying to capture when we contrast agent-relative axiologies with agent-neutral axiologies. Many important things thought to follow from agent-relative rankings (such as the possibility of consequentializing deontic constraints and special duties) actually require that the relativized rankings are different for different agents. Thus, I suggest that we make a distinction between axiologies that are agent-relative in *only* a weak sense and those that are agent-relative in a strong sense. An axiology is agent-relative in the weak sense if the evaluative rankings it contains are relativized to agents. It is agent-relative in the strong sense if the evaluative rankings are relativized to agents and different rankings are correct relative to different agents.

With this distinction in place, we can note that the strong sense of agent-relativity is the more important one when it comes to the contemporary interest in agent-relative axiologies. For example, advocates of agent-relative consequentialism have all had the strong sense in mind when developing their theories. Similar points will apply to other kinds of position-relative axiologies. Therefore, whenever I refer to any kind of position-relative axiology in this article, it is the strong sense of relativity that I have in mind.

2. Other Kinds of Relativized Rankings

2.1. Time-Relativity

Earlier we saw that deontic constraints and special duties are agent-relative rules that can only be captured by an agent-relative moral theory. However, some moral philosophers have argued that agent-relativity alone does not fully capture the deontic verdicts that follow from these rules. They have been motivated by cases like the following:

> RASKOLNIKOV: Raskolnikov's mind is gradually being overcome by homicidal thoughts. He judges correctly that (1) in the next few days he will give in to these urges and commit multiple homicides, and (2) the only option available to him that will prevent this killing spree is to commit a single homicide now (perhaps this will

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> lead to his immediate imprisonment, or satiate his homicidal urges, eradicating them from his psychology). Thus, Raskolnikov must choose between killing one innocent person now or several innocent people later.

If you accept that there is a deontic constraint on killing, what verdict should you reach in this case? Whatever Raskolnikov does, he will end up violating the constraint at some point. Nonetheless, what does the constraint require him to do right now? According to one view, he ought to kill the one now. After all, the deontic constraint gives him the aim that he does not kill and, thus, in a situation in which all available options involve him killing, he does best to at least minimize the killing that he does.

According to another view, he must *not* kill the one now. After all, if preventing another agent from murdering several innocent people cannot justify killing an innocent (p. 55) person, then why should preventing your future self from murdering several innocent people justify it? To permit the sacrifice only when it prevents your killings appears to focus excessively on your own clean hands, while neglecting the rights of the victim you sacrifice.[14]

Those who endorse the second view must interpret the constraint on killing as not just an agent-relative rule, but also a time-relative rule. In Parfit's terms, it doesn't just give different aims to different agents, it also gives agents different aims at different times. Thus, my theory-given aim is not just that I do not kill innocent people, but that I do not kill innocent people at the current moment. By contrast, those who favor the first view are interpreting the constraint as agent-relative yet time-neutral.

To incorporate the time-relative interpretation of deontic constraints into consequentialism, the consequentialist must relativize evaluative rankings not only to agents but also to times. Thus, there are not only different evaluative rankings for different agents but also different evaluative rankings for the same agent at different times. It can then be held that, for each agent, *her* killing an innocent person at a time *t* is *worse-relative-to her at t* than any number of innocent people being killed by others at any time, or by her at times other than *t*.[15] It follows that at each moment, when an agent maximizes what is *good-relative-to her* at that moment, she refrains from killing even in cases like RASKOLNIKOV where she can minimize her total killings by killing now.

Although time-relative rankings look most plausible when combined with agent-relative rankings, accepting time-relativity without agent-relativity is also possible. Broome (1991, 8) gives the example of a rule requiring us to expend all available resources saving any miners who are currently trapped in mines even if doing so will, by diminishing our resources, prevent us from saving more trapped miners in the future. Such a rule is agent-neutral (all agents are given the same aim of saving trapped miners), yet time-relative (our aims are different at different times). A time-relative, yet agent-neutral ranking could capture this rule. It would hold that at any time *t*, saving any trapped miners at *t* is better-relative-to *t* than saving more trapped miners after *t*.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

In the examples discussed so far, the relevant connection between the evaluator's position and the evaluated state has been *identical* temporal locations. However, other types of connections could also produce time-relative rankings. For example, a painful experience in my past appears to be better than a painful experience in my future. Indeed, at any particular moment, many would prefer that they had a very painful experience yesterday over being about to have a moderately painful experience.[16] This indicates that we are ranking states differently at different times. When the very painful experience is behind me and the moderately painful one is ahead of me, then I rank the latter as worse. However, if both are behind me or both are ahead of me, then I rank the former as worse. Notably, what makes the difference here is not identity in times (p. 56) but a time being before or after the temporal location of the evaluator. Although these examples are instructive, they appear to only apply to prudential evaluations we make about our own lives. For example, if I am told that my loved one either had a very painful operation yesterday or is about to have a moderately painful operation today, I prefer that she was spared yesterday's suffering and has the lessor suffering today.

2.2. Location-Relativity and World-Relativity

Recently, two further types of position-relativity have been examined by Dreier (2018). Position-relativity involves a kind of indexical reference built into rules or values. Indexicals may be indexed to persons, times, places, or worlds (e.g., "I," "now," "here," and "actually"). This suggests that in addition to agents and times, there could be relativity based on physical location (location-relativity) or possible world location (world-relativity). Are there any reasons for adopting an axiology that is position-relative in either of these ways?

Dreier (2018, 34) suggests that location-relativity could be supported by the view that the physical distance between an agent and a person needing aid is a morally significant factor that determines the strength of the obligation to aid. Because Dreier rejects this view, he rejects location-relativity. However, his discussion suggests that those like Kamm (2000; 2007, chap. 11), who endorse the view, have grounds for endorsing location-relativity. I will now argue that this is incorrect. Even if physical distance is morally significant in the duty to aid, this is best accounted for without postulating location-relativity.

The key idea behind position-relativity is that evaluators in relevantly different positions will give different rankings of states that are appropriately related to their position. This needs to be distinguished from a certain positional fact being morally significant, which can sometimes be accounted for in position-neutral terms. For example, suppose I think that it morally matters that parents help their own children. This makes agential positions morally significant but can be accounted for in agent-neutral terms (everyone shares the common aim that parents care for their children). We recognize parental duties as agent-relative because we make the further judgment that a parent must not neglect her child even if it will result in less child neglect by other parents. Thus, each parent must assess *her neglecting her child as worse* than *other parents neglecting their children*.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Applying this to the duty to aid, note that the supposed moral significance of physical distance can be accounted for in location-neutral terms. Irrespective of our location we can all recognize the value of an agent prioritizing those physically closer to her when deciding who to aid. Location-relativity is required only if we make the further judgement that an evaluator at location L_1 should assess states where an agent at L_1 prioritizes those closer to L_1 as better than states where an agent at L_2 prioritizes those closer to L_2. But this is intuitively implausible and the arguments offered by Kamm and others do not suppose anything of this sort. Given that there are no other rationales for location-relativity, we should conclude that it is implausible.

(p. 57) Let's now consider world-relativity. Suppose that we must choose whether or not to bring a large number of extra lives into existence. If we bring these extra live into existence, then our lives will be slightly worse as there will be more people to share resources with. However, the new people who exist will have very good lives that outweigh the slight loss of welfare in our lives. One might think that this example results in a kind of paradoxical situation where, whatever option we choose, its outcome is better than the other option. If we bring into existence the extra lives, we have done the best thing because, although our lives are slightly worse, this is outweighed by the good lives lived by the extra people. If we don't bring these lives into existence, then we have done the best thing because we have avoided making our lives worse and the extra welfare we might have created is irrelevant because it belongs to people who do not exist and the potential welfare of nonexistent people has no moral significance.

Dreier (2018) suggests that this verdict is best interpreted as endorsing a form of world-relativity. In this example, we are essentially choosing which of two possible worlds to bring about. If we bring about the first world (with the extra people), then, from our position as evaluators located in that world, we must rank it as better than the second world. If we bring about the second world (with no extra people), then, from our position as evaluators located in that world, we must rank it as better than the first world. Thus, different rankings are correct relative to different worlds.

One interesting feature of world-relativity is that, unlike agent and time-relativity, it appears to play no role in practical deliberation. Our choices determine which possible world is actual, so knowing that different things are best or worst relative to different worlds we can bring about cannot make one world more choice-worthy than another. Nonetheless, Dreier suggests that it is still a useful concept for the following three reasons. First, it can explain why we should feel a certain way about a choice we have made. For example, in the case we just considered it can explain why, regardless of which option we choose, we can feel good about our choice. Second, it can make sense of a strategy used by Portmore (2011) to consequentialize prohibition dilemmas. Third, it can explain Temkin's (2012) notion of "essentially comparative value" without appealing to the counterintuitive idea that "better-than" is intransitive. These are substantial claims that I cannot explore further here. However, their significance suggests that world-relativity deserves further exploration in the future.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

2.3. Patient-Relativity

In Hammerton (2016) I used the following analogy to argue that "patient-relativity" is possible. In agent-relative and time-relative rules, numerical identity between *agents* or *times* makes a difference to the deontic status of acts. For example, killing one to prevent the killing of two may be permissible when the killer of the one and the killer of the two are identical, and impermissible when they are distinct. Numerical identity between *moral patients* can also make a difference to the deontic status of acts. For example, perhaps it is wrong to break a promise to one promisee in order to keep comparable (p. 58) promises to several other promisees. However, if all the promises at stake are owed to the same promisee, then it seems permissible to break one promise to that promisee in order to keep several comparable promises to her.

I now recognize it was a mistake to describe this patient-related phenomenon as a kind of "relativity." In agent-relative and time-relative rules, identity between agents or times might make a difference to an act's deontic status. However, this is not what makes them relative rules. They are relative because they give different substantive aims to agents at different positions (positions picked out by the agent or the time). Yet there is no "patient" position that can result in different aims being given to different patient identities.[17]

We can also make this point in terms of relativized rankings. If the patient in one state is identical to the patient in another state, this may make a difference to how we ought to rank those states. However, the identity between patients is a neutral fact accessible from all evaluative positions, so it does not result in multiple correct rankings of states. By contrast, if the fact that I act in a state I am evaluating makes a difference to how I ought to rank that state, but not to how others ought to rank it, then the correct rankings of states will vary between agents.

A lesson we can draw from this is that we must be careful not to confuse *identity difference making* (identity between two things making a difference to the deontic status of acts) with *position-relativity*. Standard cases of agent and time-relativity may involve identity difference making, but this is not what makes them position-relative. A further source of confusion is that, in the cases where patient identity makes a difference to an act's deontic status, it also turns on or off certain kinds of position-relativity. For example, in the promise-keeping case, when the patients are the same, the constraint on promise breaking appears to be agent-relative yet time-neutral. And when the patients are different, the constraint appears to be both agent and time-relative.

2.4. Combining Relativities

All four kinds of position-relativity that we have reviewed so far are logically independent of one another. This means that a moral theory can, as a matter of logic, contain any combination of different kinds of position-relativity/neutrality. Nonetheless, Derek Parfit (1984) has argued that only theories that are neutral in every domain or relative in every

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

domain are plausible. He calls such theories “fully neutral” and “fully relative” and contrasts them with theories that are “incompletely neutral/relative.” Parfit says:

> This claim can appeal to the analogy between oneself and the present, or what is referred to by the words ‘I’ and ‘now’. This analogy holds only at a formal level. (p. 59) Particular times do not resemble particular people. But the word ‘I’ refers to a particular person *in the same way in which* the word ‘now’ refers to a particular time. And when each of us is deciding what to do, he is asking, ‘What should *I* do *now?*’ Given the analogy between ‘I’ and ‘now’, a theory ought to give to both the same treatment.[18]

Parfit applies his analogy only to agents and times, presumably because he was unaware of the possibility of location-relativity and world-relativity. However, the analogy he draws seems to extend to places and worlds (e.g., “What should *I* do *now*, *here*, in the *actual* world?”). Therefore, if Parfit’s argument works, what it shows is that moral theories must be either neutral with respect to agents, times, locations, and worlds, or relative with respect to them.

What should we make of Parfit’s argument? I suspect that many will find the analogy unconvincing. Why should the fact that persons, times, places, and worlds are all elements of indices that are relevant to practical decision making mean that they must all be given identical treatment? If we have independent reasons for thinking that one of these domains is relative and another is neutral, then why not accept incomplete relativity?

Some philosophers have appealed to full relativity to justify giving agent-relative deontic constraints a time-relative interpretation.[19] However, this use of Parfit’s argument seems problematic. Deontic constraints are generally defended by appeal to commonsense morality. Yet a fully relative moral theory is in conflict with fundamental features of commonsense morality. I note some of these conflicts in Hammerton (2016). Our discussion of location-relativity in section 2.2 suggests a further conflict. A fully relative moral theory must be location-relative. Yet location-relative deontic verdicts are deeply counterintuitive and there appear to be no theoretical reasons (other than the appeal to full relativity) to endorse them. On these grounds, we might conclude that if the appeal to full relativity succeeds, then it pushes us toward position-neutral theories, which, although also in conflict in commonsense morality, can at least be supported by independent theoretical considerations. Those wishing to defend agent-relative moral theories probably ought to reject the appeal to full relativity.

3. The Debate on Relativizing

In this chapter, we have noted several reasons in favor of adopting a position-relative axiology. First, it allows certain intuitive verdicts of commonsense morality to be captured from within a consequentialist framework.[20] Second, it is entailed by the combination of a fittingness analysis of value and our intuitions about what is fitting. (p. 60) And third, it allows our moral evaluations to reflect our position as particular agents, acting at a particu-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

lar time in the world we inhabit.[21] In this final section we will consider several arguments against position-relative axiologies and assess how successful they are at undermining position-relative consequentialism.

3.1. The Incoherence Argument

One argument against position-relativity questions the coherence of relativized values. Because our ordinary commonsense notion of moral value is firmly position-neutral, it is alleged that there is no conceptual space for position-relative moral values.[22] Here is Broome (1991, 8) expressing this concern:

> [The agent-relative consequentialist] claims that breaking my promise is worse *for me* than keeping it, and one might doubt that such a claim makes good sense. Certainly, there is one clear sense it could have. It could mean that breaking my promise is against my interest: that it would make me less well off than I would otherwise have been. But that is not what it is supposed to mean in the argument. The good in question is not my private wellbeing, and the argument is not appealing to my own self-interest. The good is general good, not mine; but general good evaluated from my own special position as an agent. And one might doubt that good can really be agent-relative in this way. Agent-relative teleology appeals to agent-relative good, and one might be dubious of such a concept.

And here is Schroeder (2007, 291) expressing it more stridently:

> I don't understand what "good-relative-to" talk is all about, I don't understand how it could be appealing to think that you shouldn't do something that will be worse-relative-to you. I don't even understand what that means! Until the [agent-relative consequentialists] give me some reason to think that the *good-relative-to* relation is somehow very much like the *good* property and the *good for* relation, I don't see why I should remotely find such an idea deeply compelling.

On one version of this argument, the critic claims that position-relative values are incompatible with our ordinary conception of the moral good and concludes that position-relative axiologies are not viable. However, Schroeder (2007) rightly notes that this is too quick. For, even if it lacks a commonsense precedent, the logical structure of position-relative value is clear enough to allow it to be defined by stipulation as a kind of theoretical posit. Schroeder usefully draws an analogy to "electron." We have no pretheoretic concept of an electron. Nonetheless, we are justified in employing this concept in scientific theory because it serves a useful role, allowing us to give a unified (p. 61) and elegant explanation of several empirical observations. If a useful theoretical role can also be found for position-relative value, then that could justify its postulation. Schroeder then goes on to argue that no useful theoretical role has been provided by those who defend relative values. For example, many agent-relative consequentialists appeal to the so-called Compelling Idea that it is always permissible to do what will lead to the best outcome. Schroeder argues that this can provide no independent support for postulating agent-rel-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ative values because the intuitive appeal of the Compelling Idea is based on our ordinary, agent-neutral understanding of “good” and “best.”

Two different lines of response have been offered to Schroeder’s argument. One line attempts to show that, contrary to Schroeder’s claim, a kind of position-relative value is already present in our ordinary moral thought and talk. For example, suppose that during a flood a parent decides to rescue two random children instead of rescuing her own child. Someone attempting to explain why it was morally wrong for the parent to act this way might say: “She should have recognized that, from her position as a parent, her child drowning is worse than two random strangers drowning.” This talk appears to commit the speaker to agent-relative value as “worse” is tied to the agent’s position and yet is being used in a *moral* rather than *prudential* sense (the context is one of moral blame). Therefore, insofar as people sometimes talk this way, there is room in ordinary moral discourse for position-relative values. More sophisticated versions of this line of response can be found in Dreier (2011) and Cullity (2015).

A second line of response appeals to the fittingness analysis of value. Recall from section 1.2 that this analysis of value, and our intuitions about what is fitting, imply that there are position-relative values. This means that, even if we have no pretheoretical concept of position-relative value, we can nonetheless justify it as a theoretical posit by pointing to independent reasons that support the fittingness analysis of value.[23]

Schroeder anticipates this line of response and offers a reply. He points out that, if the agent-relative consequentialist is to explain the paradoxical nature of deontic constraints, then she needs to explain why an agent must not violate a constraint even when doing so is agent-neutrally better. She can do this by supposing that, for each agent, the relative value of her not violating the constraint outweighs the neutral value of there being fewer constraint violations by others. However, Schroeder argues that if she endorses the fittingness analysis, then there is a problem. Talk of “what it is fitting to desire” requires an agent place (“what it is fitting for *x* to desire”) and so the fittingness analysis requires *all* value claims to be relativized to agents. Thus, the only way for it to accommodate neutral values is to say that a state is agent-neutrally good when it is *good-relative-to* every agent. But this makes the earlier claim about the agent-relative value of a state sometimes outweighing its agent-neutral badness incoherent. This claim is interpreted as follows: There is a state that is *better-relative-to me* but *worse-relative-to everyone* (including me). Schroeder concludes from this that basing agent-relative consequentialism on the fittingness analysis is not viable.

(p. 62) Two different counterarguments have been developed against Schroeder’s reply. Suikkanen (2009) argues that Schroeder overlooks an alternative way of accommodating agent-neutral values into the fittingness analysis. Perhaps a state is agent-neutrally good when it is *good-relative-to* the impartial spectator. Under the fittingness analysis this translates as it being fitting for the impartial spectator to desire that state. This allows for clashes between neutral and relative values to be explained as cases where a state’s *value-relative-to the impartial spectator* conflicts with its *value-relative-to me*.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

A second counterargument comes from Smith (2009). Suppose that a child narrowly escapes being crushed by a falling tree. Everyone can recognize that this state of affairs is good. What makes it good is that someone has escaped serious injury. Now suppose that the child is my child. I have an additional reason for recognizing this state as good that others lack. What makes it good is that *my* child has escaped serious injury. Smith suggests that there is an important distinction behind these two different value-making features. The former is "neutral" because it makes no indexical references to anyone and thus can be appreciated by all agents. The latter is "relative" because it contains an essential indexical reference and thus can only be appreciated by the persons picked out by this reference. This distinction between neutral and relative value-making features applies even if all value terms are indexed to agents. We saw in section 2.2 that indexing to agents is not sufficient for what I called "strong agent-relativity" because it doesn't guarantee that different rankings will be correct for different agents. However, having relative value-making features is sufficient for strong agent-relativity as it entails that there will be different rankings for different agents. Thus, we can use Smith's value-maker distinction to characterize strong agent-relativity. This provides a reply to Schroeder. In the fittingness analysis of value, value-making features can be understood as the features that make certain responses fitting. Some of these features may have the indexical reference that Smith refers to and others may lack it. Thus, we can interpret the claim that the agent-relative value of a state sometimes outweighs its agent-neutral badness as follows. Sometimes neutral value-making features count against a state, yet stronger relative value-making features count in favor of it.

The strong replies offered against Schroeder's argument suggest that the incoherence argument against position-relative axiologies is not decisive.

3.2. Theoretical Arguments

Another kind of argument against position-relative axiologies points to theoretical considerations that favor neutrality. These arguments are usually offered by agent-neutral consequentialists and, if successful, undermine not only position-relative forms of consequentialism but all nonconsequentialist theories that are relative. I will consider two such arguments here.

(p. 63) First, Derek Parfit (1984, chap. 4) argues that all agent-relative moral theories face a problem he calls "direct, collective self-defeat."[24] A theory is *directly collectively self-defeating* if and only if, when all of us successfully follow that theory, our theory-given aims are worse achieved than they would have been if none of us had successfully followed that theory. Parfit demonstrates this problem with various Prisoner Dilemma–style cases in which agents following agent-relative requirements do worse at bringing about their theory-given aims than they would have done if they had instead followed an alternative agent-neutral rule. For example, in a case he calls the "parent's dilemma," parents following an agent-relative requirement to minimize harm to their *own* child end up in a situa-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

tion where each of their children suffers more harm than he would have suffered if the parents had instead followed the agent-neutral rule of minimizing harm to *all* children.

This is a serious problem for agent-relative theories. However, Parfit shows that it can be resolved if agent-relative theories are revised so that they require agents to follow agent-neutral rules in cases of direct collective self-defeat. This solution can be adapted to agent-relative consequentialism.[25] It can escape direct collective self-defeat by fixing its axiology such that, in circumstances where direct collective self-defeat would otherwise occur, outcomes where agents cooperate to bring about that which is best relative to all are ranked higher on each agent's relative rankings than all outcomes where they do not cooperate.[26]

Although this solution is viable, it does have a bit of an ad hoc feel to it. Furthermore, it brings agent-relative theories closer to agent-neutral theories by having them adopt certain agent-neutral principles to avoid direct collective self-defeat. This leads to the obvious thought that a simpler and more elegant way to deal with this problem is to adopt a completely agent-neutral theory—a thought that Parfit has some sympathy with.

A second argument against position-relative axiologies that has similarities with Parfit's argument can be found in the work of Philip Pettit. Pettit argues that agent-relative moral theories result in a kind of "moral civil war" in which:

> two or more agents are required, and required in a self-relativized manner, each to try to advance a cause that falls particularly to them. The cause may be that of looking after certain dependents, doing one's duty in some regard, respecting the rights of those with whom one happens to deal directly, or just trying to maintain a clean record in regard to certain neutral values. (1997, 149)

(p. 64) The result is that:

> universalizing no longer means recognizing a commonly espoused goodness or rightness. All that it means is recognizing a common structure in the essentially different properties of rightness by which we are each required to steer and orientate. (1997, 149)

Pettit is right to emphasize the attractive "collective harmony" of agent-neural theories. However, the interagential conflicts that follow from agent-relative theories may be more limited than he suggests. Recall the point made in section 2.2 that I must recognize you as acting correctly when you maximize that which is *good-relative-to you* even if your act is counterproductive to my relative good. We can add to this point the observation that commonsense morality prohibits forcing, manipulating, influencing, or aiding another to do the wrong thing even in cases where doing this maximizes your relative good. Thus, suppose that you can either save your child or save mine. From my evaluative perspective it would be better if you saved my child. Nonetheless, I must recognize that it is right that you save your child (as it will maximize your relative good) and must not attempt to induce you to save mine instead. A consequentialist can capture this with the axiological

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

claim that my getting you to save my child in these unscrupulous ways is *worse-relative-to me* than you not saving my child.[27] This avoids the kind of "moral civil war" in which agents regularly interfere with each other's attempts to do the right thing. Our mutual respect for each other's need to pursue her own relative good insulates us against such conflicts. Thus, although Pettit has a fair point to make about the attractive "collective harmony" of agent-neutrality, it is not as strong as it initially appears.

Parfit and Pettit remind us that there are theoretical considerations that push us toward position-neutrality. What we have is a clash between lower-order intuitions telling us, on a case-by-case basis, that morality should respect the partiality of our situated lives and higher-order intuitions telling us that morality should move beyond our partial interests to some impartial perspective. Whether a consequentialist should endorse a position-relative axiology may ultimately depend on how she resolves this clash.[28]

References

Brook, R. 1991. "Agency and Morality." *Journal of Philosophy* 88: 190–212.

Broome, J. 1991. *Weighing Goods*. New York: Blackwell.

Brown, C. 2011. "Consequentialize This." *Ethics* 121: 749–771.

Cullity, G. 2015. "Neutral and Relative Value." In *The Oxford Handbook of Value Theory*, edited by I. Hirose and J. Olson, 96–116. Oxford: Oxford University Press.

Dougherty, T. 2013. "Agent-Neutral Deontology." *Philosophical Studies* 163: 527–537.

Dreier, J. 1993. "Structures of Normative Theories." *The Monist* 76: 22–40.

(p. 65) Dreier, J. 2011. "In Defense of Consequentializing." *Oxford Studies in Normative Ethics*, vol. 1, edited by M. Timmons, 97–119. Oxford: Oxford University Press.

Dreier, J. 2018. "World-Centered Value." In *Consequentialism, New Directions, New Problems*, edited by C. Seidel, 31–50 New York: Oxford University Press.

Forcehimes, A., and Semrau, L. 2019. "Non-Compliance Shouldn't Be Better." *Australasian Journal of Philosophy* 97: 46–56.

Garcia, J. 1986. "Evaluator Relativity and the Theory of Value." *Mind* 95: 242–245.

Hammerton, M. 2016. "Patient-Relativity in Morality." *Ethics* 127: 6–26.

Hammerton, M. 2017. "Is Agent-Neutral Deontology Possible?" *Journal of Ethics and Social Philosophy* 12: 319–324.

Hammerton, M. 2019. "Distinguishing Agent-Relativity from Agent-Neutrality." *Australasian Journal of Philosophy* 97: 239–250.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Hammerton, M. 2020. “Deontic Constraints are Maximizing Rules.” *Journal of Value Inquiry*. doi: 10.1007/s10790-020-09731-8

Hammerton, M. Forthcoming. “Agent-Relative Consequentialism and Collective Self-Defeat.” *Utilitas*. doi: 10.1017/S0953820820000096

Hurka T. 2001. *Virtue, Vice, and Value*. New York: Oxford University Press.

Hurka, T. 2003. “Moore in the Middle.” *Ethics* 113: 599–628.

Johnson, C. 2019. “The Intrapersonal Paradox of Deontology.” *Journal of Moral Philosophy* 16: 279–301.

Kamm, F. M. 1989. “Harming Some to Save Others.” *Philosophical Studies* 57: 227–260.

Kamm, F. M. 1996. *Morality, Mortality: Volume II: Rights, Duties, and Status*. New York: Oxford University Press.

Kamm, F. M. 2000. “Does Distance Morally Matter to the Duty to Rescue.” *Law and Philosophy* 19: 655–681.

Kamm, F. M. 2007. *Intricate Ethics: Rights, Responsibilities, and Permissible Harms*. Oxford: Oxford University Press.

Lopez, T., Zamzow, J., Gill, M., and Nichols, S. 2009. “Side Constraints and the Structure of Commonsense Ethics.” *Philosophical Perspectives* 23: 305–319.

Louise, J. 2004. “Relativity of Value and the Consequentialist Umbrella.” *Philosophical Quarterly* 54: 518–536.

McNaughton, D., and Rawling, P. 1991. “Agent-Relativity and the Doing-Happening Distinction.” *Philosophical Studies* 63: 167–185.

Moore, G. E. 1903. *Principia Ethica*. Cambridge: Cambridge University Press.

Nair, S. 2014. “A Fault Line in Ethical Theory.” *Philosophical Perspectives* 28: 173–200.

Parfit, D. 1984. *Reasons and Persons*. Oxford: Clarendon Press.

Pettit, P. 1997. “The Consequentialist Perspective.” In *Three Methods in Ethics*, edited by M. W. Baron, P. Pettit, and M. Slote, 92–174. Oxford: Blackwell.

Portmore, D. 2001. “Can an Act-Consequentialist Theory Be Agent Relative?” *American Philosophical Quarterly* 38: 363–377.

Portmore, D. 2007. “Consequentializing Moral Theories.” *Pacific Philosophical Quarterly* 88: 39–73.

Portmore, D. 2011. *Commonsense Consequentialism: Wherein Morality Meets Rationality*. New York: Oxford University Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Regan, D. 1983. “Against Evaluator Relativity: A Response to Sen.” *Philosophy & Public Affairs* 12: 93–112.

Scheffler, S. 1982. *The Rejection of Consequentialism*. Oxford: Clarendon Press.

(p. 66) Schroeder, M. 2007. “Teleology, Agent-Relative Value, and “Good’.” *Ethics* 116: 265–295.

Sen, A. 1982. “Rights and Agency.” *Philosophy & Public Affairs* 11: 3–39.

Sen, A. 1983. “Evaluator Relativity and Consequential Evaluation.” *Philosophy & Public Affairs* 12: 113–132.

Setiya, K. 2018. “Must Consequentialists Kill?” *The Journal of Philosophy* 115: 92–105.

Sidgwick, H. 1874. *The Methods of Ethics*. London: Macmillan.

Smith, M. 2003. “Neutral and Relative Value after Moore.” *Ethics* 113: 576–598.

Smith, M. 2009. “Two Types of Consequentialism.” *Philosophical Issues* 19: 257–272.

Suikkanen, J. 2009. “Consequentialism, Constraints and the Good-Relative-To: A Reply to Mark Schroeder.” *Journal of Ethics and Social Philosophy* 3: 1–8.

Temkin, L. 2012. *Rethinking the Good. Moral Ideas and the Nature of Practical Reasoning*. Oxford: Oxford University Press.

Notes:

([1]) When I talk of “recognizing relative values,” I am talking about recognizing the possibility of *moral* value being relative. It has long been recognized that *prudential* value (“good for”) is relative. However, this is insufficient for position-relative consequentialism (see section 1.2), even though some regard ethical egoism, which requires each agent to maximize what is *good for her*, as a kind of proto agent-relative consequentialism. The reader should also note that, although I suggest that Sen (1982) is the first consequentialist to advance a position-relative axiology, Hurka (2003, 609–612) has argued for an unorthodox interpretation of Henry Sidgwick’s *Methods of Ethics* (1874) in which Sidgwick accepts a kind of agent-relative moral good.

([2]) For example, see Scheffler (1982, 1), Pettit (1997, 129), and Brown (2011).

([3]) For a formal account, see the account I defend in Hammerton (2019), which is based on the work of McNaughton and Rawling (1991).

([4]) A couple of clarifications are helpful here. First, when we talk of “aims,” we are talking about *ultimate aims* and not *derived aims*. Given our different circumstances, I may best minimize lying by aiming to punish those who lie and you may best minimize it by aiming to reward those who are honest. However, despite our different derived aims, our ultimate aim remains the same. Second, one could interpret (3) and (4) as giving different

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

aims to different agents. For example, (3) might give me the aim that "*I* ensure that everyone is honest" and you the aim that "*you* ensure that everyone is honest." All agent-neutral rules can be reinterpreted as giving different aims in this way. However, no agent-relative rule can be reinterpreted as giving everyone the same aim (for example, there is no interpretation of (1) which gives a common aim to everyone). Thus, strictly speaking, we should say that agent-relative rules are rules that cannot be interpreted as giving the same aim to all agents, whereas agent-neutral rules are rules that can be given this interpretation.

([5]) What if everyone shares the common aim that *every agent ensures that she does not kill*? In the example we are considering, this common aim might explain why I must ensure that I do not kill rather than ensuring that you do not kill (for only the former contributes to "every agent ensuring her own nonkilling"). However, this is still inadequate as it does not give the correct verdict in a case in which I can either ensure that *I do not kill* or ensure that *you ensure that you do not kill*. So an agent-relative interpretation is still necessary. See Nair (2014) for discussion of further technical issues related to agent-neutral theories capturing deontic constraints.

([6]) For an alternative, nonstandard picture of agent-relative value that takes properties rather than states of affairs to be the bearers of value, see the appendix in Dreier (2011).

([7]) This captures an *absolute* constraint. To capture a *threshold* constraint, the consequentialist must hold that, for each agent, her killing one innocent person is *worse-relative-to her* than other agents killing *n* or fewer innocent people (where *n* is the number at which the threshold ends).

([8]) Examples of agent-relative consequentialists endorsing these kinds of axiological claims include Sen (1982, 1983), Broome (1991), Dreier (1993, 2011), Portmore (2001, 2011), Smith (2003, 2009), and Louise (2004).

([9]) In previous work (Hammerton 2016) I characterized agent-relativity in terms of numerical identity.

([10]) An upshot of this is that it may be a mistake to describe the kind of relativity we are concerned with as an "agent" kind. Mightn't a being who does not meet the full conditions for moral agency (e.g., a toddler who cannot be held fully responsible for her actions) nonetheless correctly evaluate states differently depending on whether an appropriate connection holds between herself and the content of those states? If so, then we should speak of "evaluator-relativity" rather than "agent-relativity."

([11]) On this point, see Regan (1983, 96), Broome (1991, 8), and Schroeder (2007, 272).

([12]) For example, see Garcia (1986), Smith (2003, 2009), Portmore (2007), and Cullity (2015).

([13]) See Hurka (2001, chap. 7) and Smith (2009, 268).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([14]) The first view is defended in Lopez et al. (2009) and Portmore (2011, 103–108). The second view is defended in Kamm (1989), Brook (1991), Kamm (1996, chap. 9), Louise (2004), and Johnson (2019).

([15]) This captures an absolute constraint. To capture a threshold constraint, we replace talk of "any number" with talk of "fewer than *n*," where *n* is the number at which the threshold ends.

([16]) See Parfit (1984, chap. 8).

([17]) Of course, the evaluator assessing states might be a moral patient. Furthermore, her connection to moral patients in the states she assesses might be the basis for a kind of relativity. However, this is best categorized as "evaluator-relativity" (see footnote 10) and, in any case, is not the phenomenon that I previously described as "patient-relativity."

([18]) Parfit (1984, 140).

([19]) See Louise (2004, 535) and Johnson (2019).

([20]) Portmore (2011, 103–109) and Hammerton (2020) go a step further on this point and argue that agent-relative consequentialism does *better* than non-consequentialist alternatives at accounting for our commonsense intuitions around deontic constraints.

([21]) On this last point, see especially Sen (1982, 29–30).

([22]) See especially Regan (1983) and Schroeder (2007).

([23]) See Portmore (2007) and Smith (2009).

([24]) Regan (1983, 109), and Forcehimes and Semrau (2019) also endorse versions of this argument against agent-relative consequentialism.

([25]) I show how this can be done in Hammerton (forthcoming). Parfit (1984) does not recognize the possibility of agent-relative consequentialism and so never attempts to extend his solution to it.

([26]) This only covers cases where everyone will do what they ideally ought to do. To cover other kinds of cases, we will need further axiological claims related to Parfit's principles R2 and R3. See Hammerton (forthcoming).

([27]) Sen (1983) appears to endorse something like this in response to an objection in Regan (1983).

([28]) For valuable comments on an earlier draft of this chapter, I would like to thank Ryan Cox and Daniel Nolan.

Matthew Hammerton

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Matthew Hammerton is Assistant Professor of Philosophy at Singapore Management University. He has published several articles on the structure of moral theories such as consequentialism, deontology, and virtue ethics.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Consequentialism and Action Guidingness

Frank Jackson

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.7

Abstract and Keywords

Consequentialism says that consequences settle what ought to be done. What does this imply for how we should decide, on some given occasion, what ought to be done in the light of our beliefs about the consequences of the actions available to us, our options? We explore the issues generated by the fact that typically there is substantial uncertainty about the consequences of the actions we need to choose between—we perforce must rely on the subjective probabilities of the possible outcomes of those actions. We distinguish objective "oughts" from expective "oughts" and note the complications that arise with compound actions—actions that have actions as parts.

Keywords: compound actions, consequentialism, expective oughts, Mill, objective oughts, Sidgwick, subjective probability

1. A Simple Example to Set the Scene

YOU are responsible for deciding who receives the MVP award in a local basketball competition. One question you need to address is which factors to take into account: points scored per game, assists at crucial points in the finals, blocks versus steals, and so on. We might call this the criteria question. A second question is who best satisfies the criteria that you settle on. An answer to the second question is an essential ingredient in what is needed for action; it is needed to ensure that the right person receives the award. We might call this the epistemological question, for what you need is an opinion—as good a one as you can get—about who best satisfies the criteria. Or, in preparation for what is to come, we can make the distinction in terms of "ought." In one sense, the answer to the criteria question tells you who ought to get the award—namely, whoever it is who best fulfils the criteria. In another sense, the action-guiding sense, it doesn't, because it does not tell you to whom you ought to hand the award, Jane or Mary or … For that, you need belief, hopefully reliable true belief, concerning who fulfils the criteria.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

A natural thought is that we can think of the relationship between consequentialism and action in the same general way. Consequences give us the criteria for (morally) right actions, and epistemology tells us which actions satisfy them, and thereby which actions we ought to carry out. But, as we will see and perhaps unsurprisingly, getting the details right is not entirely straightforward and there is controversy, even among those broadly sympathetic to consequentialist ways of looking at ethical issues. The troubles start from the fact that very often we do not know with any certainty what the consequences of our actions are or will be. Knowing with near certainty who satisfies some agreed-upon set of criteria for being the MVP is often easy (disagreements are usually over the criteria themselves); knowing with near certainty the consequences of our actions is often (p. 332) impossible—often we have no choice but to work with probabilities in the sense of degrees of belief or credences.

We will shortly look at a case where certainty is unattainable and note the complications and controversies that arise. There is, however, one bit of good news. Consequentialism comes in many varieties, many of which are discussed elsewhere in this handbook. As far as I can see, the key points in what follows do not require us to settle on one or another variety. We will frame the discussion that follows around relatively simple examples and steer clear of the many differences between one and another version of consequentialism, while taking for granted a broadly consequentialist way of looking at the examples, a way that (1) takes the values of the outcomes of the various actions (options) available to an agent as a key ingredient in settling what the agent ought to do, and (2) thinks of the values of these outcomes in a "downstream" way. And by the downstream way, I mean the way that holds, for example, that the reason one ought, as a rule, to keep one's promises lies in the fact that doing so has good consequences, and not in any intrinsic worth attaching as such to promise-keeping.

2. A Medical Example

Suppose that Jane Doe is treating John Doe for a serious but not life-threatening disease.[1] She has to decide between administering drug A, drug B, or drug C. These are her options in the sense of being actions she is all but certain are within her power. She believes that it is very likely that one or other of drug A or drug C would effect a complete cure but that the other one would kill John. She assigns to each a 50 percent chance of being the killer drug and a 50 percent chance of being the drug that would completely cure him. She also believes that it is very likely that giving drug B would effect a partial cure (that is, she has a high credence in its effecting a partial cure). She is (rightly) confident that there are no other possible consequences relevant for what she ought to do on this particular occasion: it isn't, for example, the case that drug B is in short supply and giving it to John would mean withholding from others in greater need. As a matter of fact, it is drug A that would effect a complete cure. Jane's opinions are all rational ones, and there is no way that she could discover that drug A is the one that would effect a com-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

plete cure or maybe no way in the time available. (What to say when these last two clauses are relaxed will occupy us shortly.)

(p. 333) What ought Jane to do in the case as described? Theorists give three different answers to this deceptively simple question. The first I will call hard-line objectivism.[2] There is just the one answer to our question: Jane ought to give drug A. What matters are the facts of the case and not the probabilities, credences, or epistemic state of our agent. Hard-line objectivists can insist that Jane would not be blameworthy if she failed to give drug A, and typically allow that if she gives drug B, as she most likely would, she should be commended.[3]

The second answer I will call inclusivism. It holds that there are two answers to our question. In one sense of "ought," the objective sense as it is often tagged, she ought to give drug A. In another sense of "ought," the expective sense as I will call it, she ought to give drug B. This second sense of "ought" is sometimes called the subjective sense because it depends in part on Jane's belief state, but, as we will see, this is an unfortunate label.

The last answer I will call hard-line epistemism. There is, it says, just the one answer to our question. Jane ought to give drug B. For what is clear in the example is that the course of action recommended by what she believes, by her credences, is giving drug B. For giving drug B is very likely to lead to good results, and although she knows that giving one or other A or C would lead to a complete cure, the 50 percent risk each has of being the killer drug means that giving either would be too risky. To say it in the language of decision theory, giving drug B has the highest expected value, the highest value when we think in terms of what happens when the values of outcomes are weighed against their probabilities. Interestingly, deontologists can agree with this answer, on the ground, as they might like to put it, that it is wrong to take unacceptable risks with other people's lives.[4]

There are, I think, serious problems with hard-line objectivism. To start with, it makes too sharp a divide between ethics and action. I think, in much company, that ethics is essentially concerned with action, with, in particular, right and wrong intentional action. It is concerned with what to do, and what we do is to act on how we take or believe things to be. That's why falling over isn't (normally) an action. It follows that we need a plausible account of what we ought to do given or relative to how we take things to be. Secondly, hard-line objectivism has trouble explaining away the intuition that there is *a* sense in which Jane ought to administer drug B. If you ask doctors what Jane ought to do in this kind of situation, they do not hesitate to say that she ought to administer drug B. Indeed, they will go further and say that if she doesn't, she will rightly be in danger of losing her licence. True, as we noted, hard-line objectivists may say that administering drug B would be commendable. But why would it be commendable? They cannot say the reason is that the action is in some sense what she ought to do. That's what they deny. What is more, the action need not be commendable. Maybe Jane goes (p. 334) ahead and administers drug B, but only because she has shares in the company that makes it. Even so, those doctors will agree that she did what she ought to do. Finally, hard-line objectivists cannot

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

explain the intuition that there is a sense in which drug B ought to be administered as one deriving from conflating an action's being what ought to be done with its being known or being likely to be what ought to be done. Administering drug B is the one action in our example that has no chance of being, by hard-line objectivists' lights, what ought to be done.[5]

Hard-line epistemism is to be preferred, I think. It rightly makes the action-guiding part of the story central: the action "good practice" manuals will tell Jane to perform, and the action she should be embarking on, is administering drug B. There is, though, a problem with hard-line epistemism; it smacks of "no speaks." We need words to describe the action of administering drug A in our example, and it seems unduly severe to refuse to let people use the word "ought" in doing so. The point is especially clear in the MVP example. Surely we should allow that there is *a* sense of "ought" on which the person who best fulfils the criteria for being MVP ought to receive the award, quite independently of opinions about who that person might be. Finally, what harm does it do to allow that there is a sense in which drug A ought to be administered, provided we are clear that there are two senses of "ought" and highlight that, in the action-guiding sense, Jane ought to administer drug B? We might also add, as I would, that in the sense of most relevance to ethics, Jane ought to give drug B, for ethics is essentially concerned with the intentional actions agents ought to perform.

This, in fact, is what is on offer in inclusivism. There is an objective sense of "ought," "ought_o," and an expective sense, "ought_e": Jane ought_o to give drug A; Jane ought_e to give drug B. The label "expective" harks back to the point noted earlier, that giving drug B is the course of action recommended by the kind of expected value way of thinking one finds in decision theory.[6] It is, I trust, by now obvious why it is potentially misleading to call "ought_e" the subjective "ought." It suggests that the sense in which Jane ought to give drug B is that it is most likely, by her lights, to be what she ought_o to do, and that is false, as we noted.

What then does inclusivism say in response to the question this chapter is about? It says that the connection between consequentialism and action guidingness is given by what agents ought_e to do. Consequentialism's answer to what an agent ought to do, in the action-guiding sense, is that it is one and the same as what an agent ought_e to do. It is as straightforward as that. Well, of course it isn't. There are a number of complications we need to discuss, and, as we will see later, we need to make a significant modification.

The first complication is really an objection in principle to the expective "ought." It is that taking what agents ought to do in the action-guiding sense to be what they ought_e to do lets agents off the hook in objectionable ways.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(p. 335) 3. Are We Being Too Kind to Mistaken Agents?

Suppose that Jane in fact gives drug B. There is a sense in which she makes a mistake. Drug A would have been better. But in saying that what she ought$_e$ to do is to give drug B, we would be giving her action a tick (check mark)—she did what she ought to do in the central, action-guiding sense—but why, runs the worry, should her failure to believe that drug A would completely cure her patient let her off the hook? It is, after all, a failure.

This line of objection can be expanded in three different ways. The first reminds us that mistakes *in ethics* do not let people off the hook. When someone refuses to prescribe contraceptives because they adhere to a (let us agree) mistaken natural law position in ethics, we do not say that, in the action-guiding sense, they do what they ought to do, on the ground that what they do accords with what they believe. Again, those who mistreat animals because they think that the pain of animals is morally irrelevant do not get excused because they are acting on what they believe. The challenge is, why excuse when the error is over a matter of fact—ignorance about drug A's being the drug that would completely cure John, in our example—when we rightly do not excuse when the error is over a matter of ethical theory? The answer is that errors in ethics are errors *in ethics*, whereas errors over matters of fact are errors over matters of fact. I emphasize that I am not saying that an error in ethics is necessarily blameworthy. It may or may not be. Despite the fact that the error is, arguably, an error over a matter which is a priori, given the huge range of opinions among intelligent people of good character about which ethical theory is correct, it would be a mistake to hold that anyone who plumps for one of the wrong theories is *automatically* blameworthy. The point, rather, is that, all the same, the error is an error in ethics.

The second way of expanding this line of objection considers a modified version of the medical case. In the new version, there is a line of investigation that Jane knows will resolve, one way or the other, whether it is drug A or drug C that is the killer drug, and thereby which would lead to a complete cure. But she cannot be bothered doing the work needed and goes ahead and administers B. It remains the case that she knows that it is very likely that one of A or C gives a complete cure and that the other kills, without knowing which is which, and that it is very likely that B partially cures. But, in the new case, it is, runs the objection, clear that Jane does the wrong thing in administering drug B. Instead, she ought$_e$ to do the work needed and then administer whichever of A or C is revealed to be the drug that delivers a complete cure. Her ignorance about drug A no longer excuses her. How, in that case, could it excuse her in the original case?

The answer to this rhetorical question is that the modification to the example means that giving drug B no longer maximizes expected value; this is why it is no longer what Jane ought$_e$ to do. There is a well-known proof due to I. J. Good—for a clear presentation, see Horwich (1982, 125*f*)—that, in cases like our modified example, carrying out the investigation and then doing whatever maximizes expected value is the (p. 336) course of action

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

that itself maximizes expected value. This means that the plausibility of the change in verdict consequent upon the change in the example in fact confirms the view that ought$_e$ is the action-guiding one.

The third way of expanding the line of objection asks about cases where Jane's epistemic state in the original version of the example is irrational. Surely, runs the thought, we should not let her off the hook if she's irrational? But if that is right, we need to make an important clarification to our account of the expective "ought." We spoke variously of probability, belief, giving such and such a chance, and credence earlier. Let us now make things a bit more precise. We can say that agents' degrees of belief or credences are their subjective probabilities, be they rational or not (provided they count as probabilities in the sense of conforming near enough to the probability calculus), whereas the degrees of belief that are rational for them in the circumstances are their epistemic probabilities. Should we specify what an agent ought$_e$ to do (according to consequentialism) as being determined by a consequentialist ranking of possible outcomes weighed against the subjective probabilities for those outcomes, or should we replace subjective probabilities by epistemic ones?

I want to bite the bullet here. Ramsey (1931) memorably said that belief is a map by which we steer. Persons act on the basis of how they take or believe things to be. Of course, sometimes one agrees to act on the basis of someone else's opinion while believing that they are wrong. Surely we have all, from time to time, said things like "OK, I'll check in the study for the missing check book but I'm all but certain that I left it at work." But in this case you are, nevertheless, acting on the basis of how you take things to be. What you believe is something like: there is some small chance that the check book is in the study; there is almost no cost to looking in the study; it would be discourteous to your partner to refuse to look in the study. The upshot is that we have no choice, it seems to me, but to specify what agents ought$_e$ to do in terms of their subjective probabilities, in terms of how they in fact take things to be. This is consistent, however, with granting that Jane would do *something* wrong in the medical example if her degrees of belief were irrational, but the wrongness will be a matter of errors in how she comes to have the degrees of belief in question. Although what she ought$_e$ to do at the time in question is, I urge, a matter of her subjective degrees of belief or probabilities at that time over the relevant outcomes, this is consistent with her having acted wrongly in coming to have those degrees of belief, in the sense of having violated one or another norm of rationality.[7]

I now turn to complications arising from actions that have parts.

4. Compound Actions

We can think of my raising both arms together as composed of my raising my left arm at the same time as I raise my right arm. We can think of a course of action over a period of (p. 337) time—playing a round of golf, say—as having early parts, parts in the middle, and later parts. Actions—actual and possible actions—with parts raise tricky questions concerning how what one ought to do regarding a part relates to what one ought to do re-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

garding the whole of which the part is a part. Much of the literature addressing these tricky questions is directed at the relationship between what $ought_o$ to be done regarding a part and what $ought_o$ to be done regarding the whole of which it is a part. However, given our remit, our discussion will be restricted to the bearing of these tricky questions on the action-guiding issue, on what $ought_e$ to be done regarding a part and its relation to what $ought_e$ to be done regarding a whole of which the part is a part. Moreover, picking up on what I said earlier, in doing this we are focusing on the questions of most importance for ethics, or so I maintain. This means that we will describe the key examples in probabilistic terms, but we will keep things simple by choosing probabilities that make it obvious which actions $ought_e$ to be done.

I will start with the synchronic case, framing the discussion in terms of an elaboration of the arms example. Suppose that it is very probable that (1) wonderful things will happen if I raise both my arms at noon; (2) acceptable things will happen if I raise neither arm at noon, and also if I raise my right arm but not my left arm at noon; (3) dreadful things will happen if I raise my left arm but not my right arm at noon. It is also very probable that if I raise my left arm at noon, I will not raise my right arm at noon. I am certain that I am completely free, at noon, to raise or not to raise either one or other of my arms, or both of my arms, and that the situation described here will not change materially in the foreseeable future (so there is no point in delaying things), and, finally, that there are no other possible consequences that need to be taken into account.

This much seems clear: I $ought_e$ to raise both my arms at noon. But what about raising my left arm at noon? On one way of looking at the issue, I $ought_e$ to raise my left arm at noon, for it is an essential part of what we have just noted is what I $ought_e$ to do at noon. On another way of looking at the issue, I $ought_e$ not to raise my left arm at noon, for it is very probable that if I did, I would not raise my right arm at noon, and it is very probable that dreadful things would happen in that eventuality.

It is far from obvious which way one should go on this question. Evidence for this is that, in my experience, many theorists think that it *is* obvious which way one should go—while differing among themselves over which way is the obviously correct way to go! Fortunately, given our remit, we can leave this issue open. Our concern is with the action-guiding question, and it is not possible for me to raise my left arm at noon without either raising or not raising my right arm at noon. The (relevant) actions available to me at noon are as follows: raising both arms, raising my left arm alone, raising my right arm alone, and raising neither arm. If we can say which out of these four options is what I $ought_e$ to do, we have done what's required of us. And of course we can—raising both arms at noon is what I $ought_e$ to do.

Now for a diachronic case.[8] I am a thirty-a-day smoker. I have finally decided that I must try and quit, and have settled on this coming Monday as the day to make a start. All (p. 338) the evidence, let us suppose, suggests that I need to choose between going cold turkey on Monday versus starting on a program of reducing to zero at the rate of one cigarette a day. I need, that is, to choose between, on the one hand, stopping smoking alto-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

gether from Monday onward, and, on the other hand, smoking twenty-nine cigarettes on Monday, twenty-eight on Tuesday, twenty-seven on Wednesday, and so on until I reach zero, and then stopping altogether. It is given that going cold turkey is best for my health and that I know this: the sooner I stop smoking entirely, the better. However, reducing to zero is still a good outcome, as it means I will stop being a smoker after twenty-nine days. Both courses of action are within my power, but, we may suppose, there is evidence from many studies of people who try to give up smoking that it is 70 percent probable that I will succeed if I set out on the reducing program and only 30 percent probable that I will succeed if I try and go cold turkey. These probabilities are my credences—I know about the studies—but we could equally suppose that they are probabilities in some other sense as far as this example goes. Finally, although succeeding in going cold turkey would be a better outcome than reducing by one cigarette a day, the difference is not nearly large enough to justify the greater risk of failure that comes along with attempting to go cold turkey.

I think it is obvious that in our example I ought$_e$ to smoke twenty-nine cigarettes on Monday, with the intention of making this the first step in the reduction program. Many agree with me, including very often smokers (who add that unfortunately the reduction program does not have the imagined success rate). But I should report that some who agree with me say that they do so only because the "ought$_e$" in question is a prudential one. But we could easily give a moral cast to the "ought$_e$" in our smoking example. Perhaps I am a leading researcher on cancer and my giving up smoking is essential to my living long enough to make a major breakthrough. In this version, many lives, not just mine, hang in the balance. It is hard to see that this would make a difference to the correct verdict.

A different line of objection is that the sense in which I ought$_e$ to smoke twenty-nine cigarettes on Monday is the sense of "ought" at work in a remark like "If you aren't going to apologize for your behavior, you ought to at least look ashamed," where it would be wrong to use Modus Ponens to conclude, on learning that no apology was forthcoming, that our miscreant did what they ought to do by looking ashamed. No they didn't; they should have apologized. But the crucial feature of this example is that apologizing is an *option*, an option that is better than looking ashamed. Whereas, in the smoking example, the relevant options on Monday are smoking zero cigarettes versus smoking twenty-nine cigarettes, and smoking twenty-nine cigarettes is better than smoking zero cigarettes. The key reason why I ought$_e$ to smoke twenty-nine cigarettes on Monday is that, out of the relevant options available to me *on Monday*, that's the best (in the expected value sense).

(p. 339) The hard question is not what I ought$_e$ to do on Monday, but what I ought$_e$ to do in the period that starts on Monday and ends twenty-nine days later, or so it seems to me. This is a fair question to ask, surely. It makes sense to ask what agents ought$_e$ to do over periods of time as well as what they ought$_e$ to do at times. But now there appears to be trouble. Out of the options available to me over the period that starts on Monday and ends twenty-nine days later is not smoking at all over that period, and that's given as the best thing for me to do over that period. It follows that what I ought$_e$ to do, starting Mon-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

day and for the following twenty-nine days, is not to smoke at all, but I can only do that if, among the things I do, is refraining from smoking on Monday.

I know that many dislike the conclusion that it is both true that what I ought$_e$ to do on Monday is to smoke twenty-nine cigarettes, and that what I ought$_e$ to do over the thirty-day period starting on Monday is not to smoke at all. They say that not only does there appear to be trouble, there is trouble. We need to rethink. Against this, I think that we should view the conclusion as a simple consequence of the fact that (1) the options available to an agent at a time differ from those available to an agent over a period of time, and (2) that what matters for what ought$_e$ to be done is a function of the options available to an agent.

The pressing question for us, given our remit, is what to say about action guidingness if what I say immediately above is correct. Where do I, the smoker, stand on the question of *what to do* on Monday—to borrow Gibbard's (2003, 5) way of capturing action directedness? One of the "ought$_e$"s points, as we might put it, towards smoking twenty-nine cigarettes on that day; the other "ought$_e$" to a course of action that has inter alia no smoking on that day.

I think that we have to acknowledge that our earlier identification of action guidingness with what an agent ought$_e$ to do was too simple. (What follows is the modification heralded when we made that identification.) This is the lesson of the smoker and like cases. In addressing the action-guiding question, we need to take into account not only whether or not some course of action is an option in the sense of being knowingly within the agent's power, but also whether or not the agent is all but certain that they will complete the option once they embark on it.[9] Let us call options with the second property, c-options. In the smoker example, the c-options on Monday are smoking twenty-nine cigarettes and smoking zero cigarettes (we may suppose that I am all but certain that I would last at least one day going cold turkey). Out of those two options, I ought$_e$ to smoke twenty-nine cigarettes on Monday, and that's the answer to what I ought to do in the action-guiding sense. Of course, if I were a different sort of person, going cold turkey for thirty days from Monday on would be a c-option for me; something that not only was something I was able to do but, in addition, something I was all but certain would ensue were I to choose on Monday to go down that route. And, in that case, it would not be true that what I ought$_e$ to do on Monday is to smoke twenty-nine cigarettes. In that (p. 340) case, I ought$_e$ to smoke zero cigarettes on Monday; what is more, over the thirty-day period starting on Monday, I should go cold turkey.

Often we (we theorists) make life easy for ourselves by discussing cases where all the options in play are c-options; we focus on examples of available actions that we can embark on, confident that we will complete them. But in cases where the options are not c-options—and the smoking case is but one among a host of cases of this kind—the answer to the action-guiding question is given by what ought$_e$ to be done out of an agent's c-options.

For more on the issues raised in this section, see Smith, Chapter 6, and Cohen and Timmerman, Chapter 7, this volume.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

5. Mill and Sidgwick

We are now in a position to make sense of some often-cited claims by important figures in the history of consequentialism.

Mill and Sidgwick were anxious to distance their versions of consequentialism (which were of a utilitarian cast) from implausible views about how to reason about and arrive at conclusions concerning what to do. They appreciated the point that often it is a mistake to seek to calculate consequences before acting: it can take too long, we aren't always very good at it, and sometimes it is just too hard, especially when the consequences concern matters and people we know little about. Mill says, "it is a misapprehension of the utilitarian mode of thought to conceive it as implying that people should fix their minds upon so wide a generality as the world, or society at large. … The multiplication of happiness is, according to the utilitarian ethics, the object of virtue: the occasions on which any person … has it in his power to do this on an extended scale … are but exceptional; and on these occasions alone is he called on to consider public utility; in every other case, private utility, the interest or happiness of some few persons, is all he has to attend to" (1861, chap. II, para. 19). And, in a famous passage, Sidgwick remarks, "it is not necessary that the end which gives the criterion of rightness should always be the end at which we consciously aim: and if experience shows that the general happiness will be more satisfactorily attained if men frequently act from other motives than pure universal philanthropy, it is obvious that these other motives are reasonably to be preferred on Utilitarian principles" (1907, 413).

Critics of consequentialism have not always been impressed by these disavowals, and you can see why. Take, for illustration, our MVP example and consider Sidgwick's remarks as they might be applied to it. Surely, runs the thought of the unimpressed, once you are clear about what it takes to be the MVP (the criterion question, in Sidgwick's terms), that's exactly what should be consciously guiding you in determining who to give the award to. It is fine, runs the thought of the unimpressed, to distinguish what makes acts morally right from how we should go about finding out which acts they are before acting, but the second is constrained by the first, as the **(p. 341)** MVP example makes clear. Drawing a distinction is one thing; showing how it meets the worry in question is another.

However, once we are clear that the action-guiding question for agents concerns what they ought$_e$ to do (out of the c-options available to them at the time of acting, as we have just been arguing; we will presuppose in what follows that the actions/options are all c-options), we can see how to make good sense of what Mill and Sidgwick say (with a bit of massaging to fit their remarks into our framework).

Let me spell this out. What matters for what agents ought$_e$ to do are the expected values of the outcomes of the various possible actions available to them. Agents ought$_e$ to adopt the action with greatest expected value, where value is thought of in consequentialist terms—say, in terms of total happiness or hedonic value. Now it is very hard to be certain of the overall consequences of one's actions. One may be certain of various, more or less

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

immediate, local consequences—for example, that giving someone thirsty a glass of water will make them happier pretty much right away—but what about the effects over the rest of the week or for all time, and what about the effects of giving the glass of water to someone else, or of giving out tea instead of water, or of writing a check to Oxfam instead of taking time to hand out water, and so on and so forth. The point is too obvious to need detailed laboring.

Now we can make good sense of what our duo say. We typically do not and cannot know exactly who will get pleasure and pain from what we do or might do, and how much pleasure and pain they will get, and how things will pan out in the long run. We may well know small fragments of the overall picture, as in our glass of water example, but that's all. Calculating, especially calculating for the world at large and for all time, isn't an option. What is an option is using degrees of belief concerning which *kinds* of actions lead to more, and which *kinds* to less pleasure and pain. As it is sometimes put, what is an option is to use various rules of thumb (Smart 1961, 30). What is more, it is a commonplace that the best way to enjoy a game of golf is to focus on the golf, not on how much one is enjoying it. Actions which have the highest probability of leading to happiness or hedonic value and so on need not be actions motivated by the desire to promote happiness or hedonic value. This is why I want my surgeon to be focused on cutting in the right places, not on how much happier I will end up being if she does. In making these last few observations, we are echoing Sidgwick's remarks. What about Mill's? We have degrees of belief concerning possible results in our vicinity and for beings we know a fair amount about, as in the thirst case. But Mill is right that it would be a mistake to cast too wide a net. We don't normally have the needed degrees of belief about what will happen far and wide.

We can in fact tweak our MVP example to make the key point. Collecting information about whom to hand the award to is a different action from handing out the award; it is a preliminary to handing out the award. What makes the example awkward for Sidgwick and Mill's purposes is the fact that, as we noted, normally collecting the information will involve focusing on the very criteria that settle who should get the award. Now let us tweak the example. Imagine that you are a local notable charged with handing out the award. You know next to nothing about basketball, and anyway the key (p. 342) numbers are not available to you. Your role is really to add gravitas and prestige to the ceremony and to make whomever receives the award feel special. What should you do in this situation? The answer is that you ought to attend to what relevant experts say about who should get the award. It is their words you need information about. Much as Sidgwick and Mill might have said.

6. The "It's Too Demanding" Objection to Consequentialism

Mackie (1977, 129) calls utilitarianism the ethics of fantasy. His claim is that no one could possibly live up to its demands. No doubt he would have said much the same for any agent neutral, maximizing version of consequentialism. Almost no one is prepared to give

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

the same weight to the interests of people they have never met as they do to, for example, the interests of their own children. When we ask what consequentialism implies by way of behavior—the action-guiding question—we get actions that almost no one will in practice carry out, or so it seems. That's the worry, and Mackie is one of many to have urged an objection of this kind to traditional versions of consequentialism.

Focusing on the expective "ought" allows traditional versions of consequentialism (by which I mean the agent-neutral, maximizing ones) to go some way toward addressing this concern. Agents know much more about the possible consequences of their actions for those they know personally, as we have just been observing; and the same goes for causes they know best—how best to help their neighbourhood school, as it might be. Agents are, moreover, much more likely to follow through on people and causes they care about and are close to, and know that focusing support in areas they are knowledgeable about typically has much better results than scattering largesse. And so on and so forth. This means that when we take into account agents' subjective probabilities of good and bad outcomes in determining what they ought to do—that is, when we ask after their expective oughts, what they ought_e to do—what we get will be actions heavily skewed toward the welfare of people, causes, and so on that are closest to them (in a wide sense of "closest," geography per se isn't to the point). Are considerations like these enough to blunt the objection? I think that Kagan (1989) gives some reason to say yes.

For more on this question, see Sobel, Chapter 11, this volume.

References

Gibbard, Allan. 2003. *Thinking How to Live*. Cambridge, MA: Harvard University Press.

Goldman (Smith), Holly S. 1977. "Dated Rightness and Moral Imperfection." *The Philosophical Review* 85: 449–487.

Horwich, Paul. 1982. *Probability and Evidence*. Cambridge: Cambridge University Press.

(p. 343) Jackson, Frank. 1991. "Decision–Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics* 101, no. 3: 461–482.

Jackson, Frank. 2014. "Procrastinate Revisited." *Pacific Philosophical Quarterly* 95: 634–647.

Jackson, Frank, and Pargetter, Robert. 1986. "Oughts, Options, and Actualism." *The Philosophical Review* XCV: 233–255.

Kagan, Shelly. 1989. *The Limits of Morality*. Oxford: Oxford University Press.

Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. Harmonsworth, UK: Penguin.

Mill, J. S. (1861). 1998. *Utilitarianism*. Edited with an introduction by Roger Crisp. New York: Oxford University Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Moore, G. E. 1912. *Ethics*. Oxford: Oxford University Press.

Portmore, Douglas W. 2011. *Commonsense Consequentialism: Wherein Morality Meets Rationality*. New York: Oxford University Press.

Portmore, Douglas W. 2015. “Review of Michael J. Zimmerman, *Ignorance and Moral Obligation*, Oxford: Oxford University Press, 2014.” *Ethics* 125, no. 4: 1236–1241.

Ramsey, F. P. 1931. “General Propositions and Causality.” Reprinted in *Philosophical Papers*, edited by D. H. Mellor, 145–163. Cambridge: Cambridge University Press, 1990.

Regan, Donald. 1980. *Utilitarianism and Cooperation*. New York: Oxford University Press.

Sidgwick, H. 1907. *Methods of Ethics*. 7th ed. Macmillan: London.

Smart, J. J. C. 1961. *An Outline of a System of Utilitarian Ethics*. Melbourne: Melbourne University Press.

Sobel, J. Howard. 1976. “Utilitarianism and Past and Future Mistakes.” *Nous* 10: 195–219.

Willard, E. 2005. “Monkeys, Typewriters, and Objective Consequentialism.” *Ratio* 18: 252–360.

Zimmerman, Michael J. 2008. *Living with Uncertainty*. Cambridge: Cambridge University Press.

Notes:

([1]) There are many examples of the kind I am about to describe in the ethical literature; see, e.g., Regan (1980, 264*f*) and Jackson (1991). However, stockbrokers have long been familiar with cases of this kind. The shares they advise buying are often ones they are certain will not deliver the best return. They know that the shares that will deliver the best return are one or another of the highly speculative ones; they only wish they knew which one.

([2]) Some hard-line objectivists are implicitly so; they presume (to put matters in terms of our example) that giving drug A is the answer to what Jane ought to do. But, e.g., Smart (1961, 20*f*) is explicit on the point, as is Moore (1912).

([3]) See, e.g., Moore (1912, 81–82).

([4]) See Portmore (2015).

([5]) Objections of the kind given here to hard-line objectivism have a history. See, e.g., Willard (2005), Jackson (1991), and, for a detailed discussion, Zimmerman (2008).

([6]) It is often called the prospective “ought”; see, e.g., Regan (1980). Calling it the epistemic “ought” would invite confusion with the “ought” in “The cool change ought to arrive soon.”

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([7]) For dissent on this point see, e.g., Zimmerman (2008, 34–35), but it is dissent against a background of much agreement.

([8]) There are many examples of this general kind in the literature, see, e.g., Goldman (Smith) (1977), Sobel (1976), and Jackson and Pargetter (1986). As noted in the text, often the examples are framed (unfortunately, as I now think) in terms of ought_o; sometimes this is explicit, sometimes it is implicit. For more on the differences consequent on framing matters in terms of "ought_e" instead of "ought_o," see Jackson (2014).

([9]) In what follows, I am influenced by Portmore (2011). I am not suggesting he would agree with the use I make of what he says.

Frank Jackson

Frank Jackson is Emeritus Professor at The Australian National University. He works in the philosophy of mind, ethics, and the philosophy of language. His books include Conditionals (Blackwell, 1987), From Metaphysics to Ethics (Oxford, 1998), and Language, Names, and Information (Wiley-Blackwell, 2010).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Consequentialism, Ignorance, and Uncertainty

Krister Bykvist

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.8

Abstract and Keywords

Act consequentialism provides an answer to the question of what one ought to do, no matter which situation one is in. The problem, however, is that this answer is rarely if ever known by the agent herself. Ordinary agents do not know all the consequences of their actions, nor do they know how to assess all possible consequences. This lack of both empirical and evaluative knowledge means that ordinary agents often, or always, do not know what act consequentialism tells them to do. This "ignorance challenge," as we might call it, is often seen as one of the main challenges for act consequentialism. This chapter discusses the main responses to this challenge. It also asks whether this challenge is unique to consequentialism.

Keywords: consequentialism, ignorance, uncertainty, action-guiding, rationality

1. Introduction

ONE virtue of traditional act consequentialism is its theoretical simplicity: it simply tells you to maximize value. More exactly, it says that you ought to perform the action that has better consequences than that of any alternative action. Furthermore, given a ranking of outcomes, act consequentialism has clear implications for every possible choice situation. It thus provides an answer to the question of what one ought to do, no matter which situation one is in. The problem, however, is that this answer is rarely (if ever) known by the agent herself. Ordinary agents do not know all the consequences of their actions, nor do they know how to assess all possible consequences. This lack of both empirical and evaluative knowledge means that ordinary agents often, or always, do not know what act consequentialism tells them to do. This "ignorance challenge," as we might call it, is often seen as one of the main challenges for act consequentialism.

Some understand the challenge in terms of *action-guidingness*: a plausible moral theory must be action-guiding, but how can we be guided by consequentialism if we can never know what it tells us to do? Others understand the challenge in terms of providing the *wrong prescriptions*: it does not seem plausible to say that we ought to do what *in fact*

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

maximizes value when we have good reason to think that this action does not maximize value; instead, what we ought to do should be more closely tied to our epistemically impoverished perspectives.

In this chapter, I shall critically examine some replies to the ignorance challenge for consequentialism. I shall begin by discussing possible replies to the challenge of lacking normatively relevant empirical knowledge. After that I will move on to discuss the challenge of lacking evaluative knowledge. As will emerge, there are four main approaches (p. 311) to the ignorance challenge, both in its empirical and evaluative guise (by "ought" I mean "morally ought" here and in the rest of chapter):

(a) *Just bite the bullet*: accept that we ought to do what in fact maximizes value and claim that there is nothing more of normative significance to be said; it is just an unfortunate aspect of our predicament that we never know what we ought to do.
(b) *Invoke different senses of "ought"*: claim that there is one sense of "ought," the *objective* sense, according to which we ought to do what in fact maximizes value, but add that there are other senses of "ought," according to which this is not true. On these senses of "ought," what we ought to do is tied much closer to the epistemically impoverished situation of the ordinary agent. For example, on one such reading, it is true that we ought to do what we think (or has good reason to think) maximizes value.
(c) *Change the ought-makers*: claim that there is only one (relevant) sense of ought, but we should not think that what makes an action something we ought to do is that it in fact maximizes value; what makes an action something we ought to do is instead some *mental* fact about the agent, for example, the fact that he thinks that the action maximizes value, or some *epistemic* fact about the agent, for example, the fact that he has good reason to think that the action maximizes value.
(d) *Invoke rationality*: claim that even if we do not know what we ought to do, because we do not know what will in fact maximize value, there can still be a rational action for us to perform given our ignorance. For example, as I will show in sections 4 and 6, if you are a morally conscientious agent who cares about the goodness and badness of outcomes, there is a rational action for you to perform given your preferences over possible outcomes and your credences in those outcomes.

The "just bite the bullet" approach is inconsistent with any of the other approaches (assuming rationality is normative), but some of the other approaches can be combined. For example, as I will show in sections 4 and 6, one can invoke rationality and define a sense of moral "ought" partly in terms of it.

Finally, I will compare consequentialism to other moral theories and ask whether is it only consequentialism that has to grapple with an ignorance challenge. If it is a common problem for all theories, much of the critical edge of the ignorance challenge to consequentialism is lost.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

2. Background Assumptions

I shall focus on *maximizing act consequentialism*, since this is usually picked out as a clear example of a consequentialist theory that faces the challenge of ignorance head on (p. 312) (for other forms of consequentialism, see Greaves, Chapter 22, this volume, and Hooker, Chapter 23, this volume).

I also make some metaphysical assumptions. I assume a noncausal counterfactual analysis of outcomes, according to which the outcome of an action is the whole possible world that would be actual, if the action were performed (see Dorsey, Chapter 5, this volume). This analysis assumes in turn *counterfactual determinism*: for each action, there is a determinate possible world that would be actual, if the action were performed. These metaphysical assumptions are of course very contentious, but they simplify the discussion considerably, and much of what I say can be applied to act-consequentialist theories that relax these assumptions. For example, a consequentialist theory that acknowledges *objectively chancy* outcomes will also face the challenge of ignorance since agents do not normally know the objective chances of the outcomes of their available actions. The approaches listed earlier can also be applied to the case of ignorance of objective chances.

Finally, I shall make two simplifying assumption when I discuss uncertainty. First, when I discuss empirical uncertainty, I shall assume evaluative certainty, and when I discuss evaluative uncertainty, I will assume empirical certainty. Second, I shall assume that the agent knows which actions are the relevant alternative actions in the situation. As Portmore (2011) points out, this is not always true (see also Smith, Chapter 6, this volume). Obviously, we can face both empirical and evaluative uncertainty in the same situation, but to deal with both uncertainties simultaneously would add too much complexity for a short handbook chapter like this. For the same reason, I have bracketed the issue about uncertainty about the set of alternative actions.[1]

3. Empirical Uncertainty

It is clear that ordinary agents do not know all the morally relevant empirical facts about the consequences of their actions. This is especially clear if consequences are identified with whole possible worlds. No ordinary agent knows all the morally relevant facts of the past, the present, and the future, including the far future, that would be actualized if he performed a certain action. For example, no one knows all the morally relevant facts—past, present, and future—about the consequences of global warming if we do not do anything about it, nor do we know all the relevant facts—past, present, and future—about the consequences of various ways of counteracting global warming. But this means that we do not know what we ought to do to counteract global warming, assuming that consequentialism is true.

(p. 313) There are many smaller-scale versions of this problem, too. No ordinary doctor is completely certain about the morally relevant consequences of the possible treatment op-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

tions at hand. But, again, this means that no ordinary doctor can know which treatment he ought to choose, assuming consequentialism is true.

The "just bite the bullet" response to these cases is just to say that we ought to counteract global warming by choosing the act that in fact maximizes value, and the doctor ought to choose the treatment that in fact maximizes value. The "just bite the bullet" approach adds that there is nothing more of normative significance to say about this. It is just a tragic fact of life that we often, or always, are ignorant about what we ought to do.

Not surprisingly, this approach has not had many proponents. Even a staunch objective act consequentialist like G. E. Moore thinks there is something more to be said, something that has normative significance. Suppose that you try to act rightly, taking into account all available evidence of which action has the best consequences, but still fail to act rightly because you lack knowledge of which action would have the best outcome. Moore would then say that you do not deserve blame, because you have a good excuse: "I tried my best to act rightly but I failed because I could not have known which action had the best consequences" (Moore 1912, 82). This reply has limited significance, however.

First, to say that we ought to maximize value, even though you can never know which action does that, and then just leave it at that seems to paralyze moral decision-making. Surely, we need to be able to say something more about what we should do in this kind of case. Indeed, Moore himself seems to suggest as much when he talks favorably about an agent who acts on available evidence about which action has best consequences. The implicit thought seems to be that this is what the agent *should* do, in a sense that is different from the objective sense.

Second, there are cases where the agent knows that an action will not have the best consequences, but the action is still something the agent should go for. One such case is the famous Jackson case.[2] A doctor must decide on the correct treatment for a patient who has a serious skin complaint. Careful consideration of the literature has led her to the following conclusions. B will relieve the condition but will not completely cure it. One of A and C will completely cure the skin condition; the other will kill the patient, but there is no way she can tell which of the two is the perfect cure and which is the killer.

To make the structure of the case clearer, let's put it in a table, where it is made explicit that the agent's evidence divides equally between the S1 and S2 and what the values of the outcomes are (measured on a cardinal scale, so we can compare value differences; see Table 16.1).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Table 16.1		
Actions	States of Nature	
	S1 (p = 0.5)	S2 (p = 0.5)
A	Complete cure	Death
B	Partial cure	Partial cure
C	Death	Complete cure

The intuitively right option is B even though this is guaranteed to be suboptimal in terms of the patient's well-being (and total value, assuming other things are equal (p. 314) between the outcomes of A, B, and C). The objective consequentialist, on the other hand, must say that the right option is either A or C. But both A and C seem wrong since, for each action, the possible gain of performing it does not compensate for the possible loss. To perform either action seems reckless.

Note that here the objective consequentialist cannot soften the blow by distinguishing between wrongness and blameworthiness and say that even though the agent would act wrongly in doing B, he does not deserve blame, for the agent *knew* at the time that doing B would be suboptimal in terms of the patient's well-being (and total value) and thus wrong, according to an objective consequentialism (Zimmerman 2008). The objectivist can't therefore claim that the agent is blameless because he acted in good faith, falsely but justifiably believing that he was doing the right thing. It is therefore unclear how the agent could avoid blame.

One option for the consequentialist is to replace objective consequentialism with a more subjective version of consequentialism, and thus change the ought-makers:

> *Subjective expected value consequentialism*
>
> An action ought to be done if and only if (and because) it maximizes subjective expected value.[3]

To identify the action that maximizes subjective expected value you do the following.

(1) List the possible outcomes of an action.
(2) For each possible outcome, ask yourself how probable you think it is that the action will have that outcome (i.e., how strongly you believe that the action will have that outcome).
(3) For each outcome, multiply the subjective probability of the outcome with the value you think it has in terms of total well-being.
(4) Sum these products, and you have the subjective expected total well-being of the action.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(5) Repeat this procedure for all alternative actions.
(6) Choose the action that has the highest subjective expected value.

(p. 315) This theory will provide the right answer in the Jackson case. This theory will give us prescriptions in many other cases of uncertainty too. For instance, suppose you are approaching the blind intersection and you are pretty sure (probability 0.7) that nothing will happen if you just drive through it without stopping. However, your remaining credence (probability 0.3) is for the hypothesis that you will cause a serious accident if you do not stop. Then subjective expected value consequentialism will tell you to stop (assuming that driving through does not provide any significant gain).

But this theory seems to be all too subjective, since it is defined in terms of what the agent happens to *think* is probable and valuable. What if the agent's probabilities assessments are seriously misinformed? Suppose that a doctor prescribes a certain medicine for a minor skin complaint in the belief that it will be an effective and harmless cure. His belief is not well grounded; he only has a hunch that this is the right medicine. Now, if all evidence available to him suggests that the medicine will very likely have only harmful effects—all the facts about the medicine are there in a book on his desk—would we still want to say that it is right for the doctor to prescribe the medicine? The obvious option is to reformulate the consequentialism so that it instead takes into account the *epistemic* probabilities of the agent, roughly, the probabilities he has good epistemic reason to assign to the outcomes of his options. This epistemic view about expected value will take care of the sloppy doctor, for even if he did not in fact assign a high probability to the outcome that the medicine will have harmful effects, he still had good epistemic reason to do so.

It is not clear that going for epistemic probabilities is a sufficient fix, however. Our available evidence can be seriously misleading. For example, one can have misleading evidence that suggests that the only way to avoid some disaster is to torture some innocent people. Is it really right to say that we ought to torture the innocent people? Or to take another classical example, think about Oedipus, who killed his father and made love to his mother (Oddie and Menzies 1992). Judging by our common-sense views on patricide and incest, he seems to have acted wrongly on both counts. But what makes the story so tragic is that he did wrong even though he tried his best in the light of available evidence to avoid doing these wrongs. The objectivist consequentialist can accept these verdicts, assuming, which is likely, that killing his father and making love to his mother would not have the best consequences. Furthermore, note that if the epistemic view is accepted, it is never true that what makes an act right is that it brings about the best consequences, not even if when it is known that the act has the best consequences. What makes an action right is always that it maximizes expected value. The problem in a nutshell is that the epistemic view seems to provide neither the correct moral prescriptions nor the correct ought-makers in all cases.

Perhaps we can live with the epistemic view's unintuitive moral prescriptions and ought-makers if the principle is action-guiding, but is it? Well, that depends on what exactly one

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

means by action-guidingness (see Jackson, Chapter 17, this volume). If by an action-guiding principle, we mean a principle that we can *consciously* apply in the sense that we can be motivated to perform the action that we *believe* has an ought-making feature according to the principle, then the epistemic view is action-guiding but so is (p. 316) objective consequentialism, since we can be motivated to perform the action that we believe maximizes value. If by an action-guiding principle, we mean a principle we can *successfully and knowingly* apply, then objective consequentialism is not action-guiding, except for near-omniscient agents.[4] But it is questionable whether the epistemic view is action-guiding in this sense for all ordinary agents in all cases. After all, to identify the ought-makers on this view, you need to identify the possible outcomes of an action; assign the epistemic probability to each outcome; then for each outcome, multiply the epistemic probability of the outcome with the total value you think it has, sum these products; repeat this procedure for all alternative actions; and, finally, choose the action that has the highest expected value. This is a daunting and time-consuming task, and ordinary agents are very likely to make mistakes, except in those rare cases where you have very few actions and very few possible outcomes and enough time to do the necessary calculations (Feldman 2006).

So, even if the epistemic view is more action-guiding than objective consequentialism, since ordinary agents can at least sometimes apply it successfully and knowingly, this is not much comfort, since in more complex cases the epistemic view cannot be applied in this way. This also suggests that action-guidingness should not be given too much weight in the choice between theories. It is a wild-goose chase to look for a theory that can always be successfully and knowingly applied by all ordinary agents (Zimmerman 2014, 90–112). Only ridiculously simple consequentialist theories, such as the theory that tells you to do what you have a hunch will maximize value, will be action-guiding in this strong sense. But such theories are clearly unacceptable.

Is there a version of consequentialism that can avoid the problems with providing intuitively wrong prescriptions and ought-makers and still give plausible verdicts in the Jackson case as well as providing some guide to action?

The “invoke different senses of ‘ought’” approach seems to be a good candidate. On this account, one could “have one’s cake and eat it too” in cases of uncertainty. What makes an action objectively right is always the fact that it maximizes value. But what makes an act right in some other nonobjective sense can involve the agent’s evidence about probabilities and values. For example, one could argue that there is a nonobjective sense of “ought,” according to which one ought to maximize expected value. This would mean that one can both say that, in the Jackson case, the doctor ought to completely cure the patient and she ought to choose the partial cure. This is coherent, since the first ought is objective and the second is nonobjective. Furthermore, one can agree with Moore that there can be blameless wrongdoing when the wrongness is objective. But when you act wrongly in some nonobjective sense you are blameworthy.

Now, the crucial question for this account is how to characterize the nonobjective senses of ought. More specifically, how are they related to the objective sense?

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(p. 317) One popular idea is to define objective ought in terms of nonobjective ones, for example, what you objectively ought to do could be defined as what you nonobjectively ought to do in a situation in which you know all the morally relevant facts, including the consequences of your actions and their value.

This account faces problems, however, since the values of consequences can depend on your ignorance in different ways. Going to your friend tonight would be best thing to do, since there will be a surprise party for you there. However, if you were to know that there is a surprise party there, going to your friend would no longer maximize value—the party will no longer be a surprise for you and this will make it less enjoyable for you and everyone else at the party (Schroeder 2007).

Another option is to define nonobjective ought in terms of objective ought. But exactly how should this be done? There are many options. Here are a few:

> You nonobjectively ought to perform an action = df. you believe that it objectively ought to be done.
>
> You nonobjectively ought to perform an action = df. you think it is more likely to be what you objectively ought to do.
>
> You nonobjectively ought to perform an action = df. you are justified in thinking that it objectively ought to be done.
>
> You nonobjectively ought to perform an action = df. the action would maximize expected deontic value, where the deontic value of an action is a measure of how close the action is to be what objectively ought to be done, and the closeness is defined as the difference in the value of their outcomes.(Oddie and Menzies 1992; Portmore 2011, 19)

The list could go on. Are they all genuine oughts of equal importance or is one of them more privileged? Note that only the last one seems to give the intuitively right verdict in the Jackson case. Note also that none of them is action-guiding in the strong sense that any ordinary agent can always successfully and knowingly apply it in all cases. Obviously, we can be mistaken about epistemic facts. But facts about what we believe are not always transparent to us either, as any psychotherapist can bear witness to.

4. The Rationality Account Applied to Empirical Uncertainty

Instead of ironing out all the wrinkles with the “invoke different senses of ‘ought’” account, I would like instead to focus on the “invoke rationality” approach and see how far this will take us. The idea is to stick to the objective version of consequentialism and bring in decision-theoretical rationality to deal with uncertainty (Bykvist 2009b; Graham

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

2010; Oddie and Menzies 1992; Smith 2006, 2009). One advantage of this approach is that it invokes a notion of rationality that should be acceptable to all sides of the debate.

I shall first show how we can knowingly do objective wrong in the Jackson case and get away with it in the sense that we are still epistemically, morally, and rationally sensible. The argument can be divided into three steps (Bykvist 2011).

Step 1

Suppose that your evidential probability divides equally between two empirical hypothesis H1 and H2, and that you have two alternative actions, A and B, and the possible outcomes have the values as depicted in Table 16.2.

Table 16.2

Actions	Empirical Hypotheses	
	H1 (p = 0.5)	H2 (p = 0.5)
A	10 (right)	–100 (wrong)
B	10 (right)	10 (right)

Since we are assuming that the agent is *morally* conscientious, he will prefer better outcomes to worse and be indifferent between outcomes that have the same value. More exactly, the agent will prefer outcome (B, H2) to (A, H2) and be indifferent between (A, H1) and (B, H1). Since we are also assuming that the agent is *epistemically* sensible, he will proportion his beliefs to the evidence given in the table; that is, he will have credence 0.5 in H1 and credence 0.5 in H2. Now, if he is also minimally *rational*, he will prefer option B to A, given his beliefs and preferences, since he would not prefer risking value without any possible gain. More generally, as a minimally rational person he prefers bringing about the prospect (x, 0.5, y) rather than the prospect (z, 0.5, u), if he prefers x to z and is indifferent between y and u.

Note that this shows that what is rational to do need not coincide with what is objectively morally right. Suppose that H2 is true, so that A is objectively wrong. Still, A is rational given the agent's beliefs and preferences.

Step 2

In this step we alter the situation slightly by stipulating that, if H1 were true, the outcome of B would be slightly worse than the outcome of A (see Table 16.3).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Table 16.3		
Actions	Empirical Hypotheses	
	H1 (p = 0.5)	H2 (p = 0.5)
A	10 (right)	–100 (wrong)
B	6 (wrong)	10 (right)

As a morally conscientious agent the agent does not just prefer better to worse outcomes, it is also true of him that the strength of his preference for an action over another is proportionate to the difference in value between the outcome of the two actions, so that the greater value difference, the stronger preference.

(p. 319) Of course, this means that the agent is not just concerned with acting rightly, that is, doing what is best, and avoiding acting wrongly, that is, doing what is suboptimal. But this seems right, if we think about what a morally conscientious agent would be according to consequentialism. If consequentialism is true, it would be wrong for the agent not to be sensitive to the fact that the value difference between (H1, A) and (H1, B) is much less than that between (H2, B) and (H2, A).

This means that, given his beliefs and preferences, it is still rational for him to prefer B to A, since the possible gain in outcome-value does not compensate for the possible loss in outcome-value. More generally, as a rational person, he prefers bringing about the prospect (x, 0.5, y) rather than (z, 0.5, u), if his preference for x over z is much stronger than his preference for u over y.

Step 3

Here we introduce an extra option C, associated with a prospect that is just a permutation of A's (see Table 16.4).

(p. 322) Table 16.4		
Actions	Empirical Hypotheses	
	H1 (p = 0.5)	H2 (p = 0.5)
A	10 (right)	–100 (wrong)
B	6 (wrong)	6 (wrong)
C	–100 (wrong)	10 (right)

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

In this situation, it is still rational for the agent to prefer B to A, given his beliefs and preferences, since C's prospect is just a permutation of A's, and the agent should thus be indifferent between C's prospect and A's. More generally, a rational person is indifferent between (x, 0.5, y) and (y, 0.5, x). But then we get the result that for an epistemically, morally, and rationally sensible agent, it is rational to prefer B to all alternatives in a Jackson case. In other words, it is rational for him to prefer doing something that he knows is morally wrong. Note that in this case it is *impossible* to do act rationally and at the same time do what is objectively morally right.

(p. 320) Of course, the rationality principles invoked here are far from a complete theory of rationality. They are entailed by the rationality principle that tells us to maximize expected utility, but they do not mandate it. Even with these rough contours of the rationality theory, it is clear how it is supposed to work in other cases of uncertainty, such as the risky intersection. What makes an action right in these situations is that it in fact maximizes value. However, even in these cases there is a rational action, given the agent's credence and preferences. It is true that it is not always easy to identify one's credences and preferences, and thereby be successfully and knowingly guided by them in one's decisions, but in this respect the account is not worse than the epistemic view, which also required introspective knowledge of similar sorts.[5] Finally, there is a clear sense in which this account gives an action-guiding role to the objective version of consequentialism: the agent's ultimate concern are the values that are objective ought-makers.

However, the "invoke rationality" account has some objections to answer (Bykvist 2011).

(a) *Morally conscientious wrongdoing*. Zimmerman would object that I have given more of a caricature than a plausible characterization of a morally conscientious person, since he insists that no conscientious person who adopts an objectivist theory could prefer option B in the Jackson case, since he knows that it is wrong and has a suboptimal outcome.

In reply, note first that we should all agree that no conscientious moral agent would knowingly do wrong *if he knows of some other options available to him that it would be right*. But the cases we are now considering are not cases where the agent has knowledge of what is the right option. In these cases, it would be morally insensitive to care only about doing right and avoiding doing wrong, since the moral theories under consideration discriminate between the outcome-values of different wrongdoings. A sensible moral agent should be willing to sacrifice a possible loss in outcome-value for the sake of a sufficiently great gain in outcome-value.

(b) *Blameless wrongdoing?* If knowingly doing wrong implies blameworthiness, I have a problem, since I will then have to say that the agent in preferring or choosing B deserves blame. But the link between wrongness and blameworthiness is not this simple. There are many cases where knowingly doing wrong does not merit blame, namely, those cases in which you have a valid excuse. In the Jackson cases under consideration, it seems obvious that you have an excuse: you can't tell which action is morally (p. 321) right. It would thus be unfair to blame you for knowingly doing wrong when you are being sensitive to degrees of outcome-value. On the contrary, you would be blameworthy if you only cared

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

about avoiding wrongdoing and ignored morally relevant information about the differences in the outcome-values of different wrongdoings.

(c) *Lack of moral significance*. One could complain that the rationality account does not make empirical uncertainty a *sufficiently moral* issue for the following two reasons. First, since we are talking only about rational action here, we seem not to be able to say that there is something *morally* problematic about risking a really bad outcome for some small gain in outcome-value. For example, in the Jackson case, if you choose A, and it does in fact give a complete cure, we can say only that you acted irrationally. But choosing A seems morally problematic, even though it happens to give a complete cure.

Second, since we are talking about what is rational to do, *given* one's preferences and beliefs, these prescriptions will only apply to agents who have preferences concerning the values of outcomes (Zimmerman 2014, 46). How can we say that this account takes empirical uncertainty sufficiently seriously, if it has nothing to say to agents who lack such preferences?

In reply, one could point out the rationality account can say that there is something *morally* problematic about your action of taking a great risk, even if you do not care about it. To take such a risk is to do something that a morally conscientious and rational person would not do in your situation, if she shared your credences about empirical matters. Hence, it is to do something that an agent, who has a certain *moral virtue*, would not do in your situation. But to do something that an agent with a certain virtue would not do is to act in a less than fully *virtuous* way with respect to this virtue.[6] The rationality account could thus be seen as providing a kind of (external) virtue assessment of actions, which is applicable even to agents who lack the preferences a virtuous agent would have.[7] More exactly, your action is virtuous in this respect just in case it would be done by a morally conscientious and rational agent, who shared your empirical credences. This account also gives us a kind of *moral ought*: what one ought to do in order to act virtuously in this respect.

This reply also shows that the rationality account and the "invoke different senses of 'ought'" account need not be seen as competitors, since what we have just done is to define a nonobjective sense of "ought": what one ought to do in order to act virtuously in a certain respect, which depends in part on one's empirical evidence rather than the empirical facts.

5. Evaluative Uncertainty

So far we have focused on cases of empirical uncertainty. What has received much less attention in the debate is the problem of *evaluative* uncertainty: what should you do when you are not sure about your value judgements. For instance, what should you do if you think that both human and animal lives are valuable but not sure how valuable humans are as compared to animals? What should you do if you take seriously the risk of human extinction but are not sure how bad this would be? Some philosophers have argued that

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

the problem of evaluative uncertainty can only be solved if the consequentialist theory is radically revised so that it takes into account the agent's evidence for various evaluative hypotheses.

According to one version of this approach, which we may call *Value Vacillator Consequentialism* (VVC), an action is judged by the probable value of its probable outcomes. More exactly, the value of action is a function of (a) the probabilities of its possible outcomes, (b) the probabilities of evaluative hypotheses about the values of outcomes, and (c) the intrinsic values of outcomes proposed by these hypotheses. VVC is thus a version of the "change the ought-maker" approach, now applied to evaluative uncertainty.

According to Michael Zimmerman's version of VVC—which I will focus on in the following since it is by far the most worked-out theory to date—an action ought to be performed iff it is has the greatest *expectable* value, and the expectable value of $A = \Sigma_i$ prob $(O_i/A) \times \Sigma_j$ prob $(O_i$ has $V_j) \times V_j$, where O_i is a possible outcome of A, and V_j a possible value (Zimmerman 2008, 39).

Zimmerman assumes that the relevant probabilities are epistemic, which means that the higher epistemic probability of an outcome, the higher degree of belief is supported by the agent's total available evidence.

Here is Zimmerman's own example of how his VVC will play out in a case of uncertainty about the value of nonhuman versus the value of animal well-being (Zimmerman 2008, 19). Suppose that you are certain that giving a certain pill to John will cure him partially, and you are also certain that giving this pill to Jane will cure her completely. You know that John is a human being and Jane is a hamster, and you know how your actions would affect their well-being, but you are not sure how *valuable* animal welfare is compared to human welfare. Your evidential probability divides equally between *impartialism*, according to which the value of a partial cure of John is 100 and the value of a full cure of Jane is 120, and *speciesism*, according to which the value of a partial cure of John is 100 and the value of a full cure of Jane is 20 (see Table 16.5).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Table 16.5			
Actions	Evaluative Hypotheses		Expectable Value
	Impartial (p = 0.5)	Speciesist (p = 0.5)	
Give pill to John (human)	100	100	100 (100 × 0.5) + (100 × 0.5)
Give pill to Jane (hamster)	120	20	70 (120 × 0.5) + (20 × 0.5)

Zimmerman's theory would thus tell us to give the medicine to John, since this has the greatest expectable value. By contrast, traditional consequentialism will say that we ought to give the pill to Jane, if the impartial hypothesis is true, and that we ought to give the pill to John, if the speciesist hypothesis is true.

VVC has some objections to answer, however.

(p. 323)

(a) "*The wrong right-makers.*" Traditional consequentialism identifies the right-makers of actions with facts about the intrinsic value of outcomes (or, more fundamentally, with facts about the value-makers of these outcomes, i.e., the natural features that makes outcomes good or bad). In contrast, VVC identifies the right-makers of actions with facts about the agent's *evidence* for outcomes and his evidence for their intrinsic value. But this means that facts about the actual intrinsic values of outcomes will never be right-makers (and facts about the natural features that make outcomes good will never be the more fundamental right-makers). Indeed, not even in a case of full knowledge of the intrinsic values of the outcomes will the actual intrinsic values of outcomes be the right-makers of actions. Suppose, for instance, that someone is kicking a cat for fun, and that he knows that it is bad for the cat to suffer. We would like to say that it is the fact that it is bad for the cat to suffer that in part makes the action wrong. The proponent of VVC, however, has to say that even in this case what makes the action wrong is not that the fact that it is bad for the cat to suffer but that the fact that the agent has *evidence* for the hypothesis that the cat's suffering is bad. One could therefore object that VVC does not take actual intrinsic values (or actual intrinsic value-makers) seriously (Bykvist 2014). Indeed, if one concedes that the actual values of outcomes *never* make an action right, how could one honestly claim to be a consequentialist?

(b) *Morality sanctions serious evildoing.* Suppose that you are faced with the choice between torturing animals for the sheer fun of it or refraining from doing so, and

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> suppose, not surprisingly, that it would actually be best to refrain from doing so, but that, due to serious deficiencies in your evaluative evidence, it would maximize expectable value to torture animals and enjoy doing it. Perhaps you have evidence that your own well-being is much more valuable than that of animals. Are we really willing to say that under such circumstances you morally ought to torture the animals and enjoy doing it (Bykvist 2014; 2018)?

Zimmerman is willing to bite this last bullet (Zimmerman 2008, 75). He claims that traditional consequentialists are pretty much in the same boat, if they accept cases of *blameless wrongdoing*, which they typically do. For instance, if an agent's evidence was seriously (p. 324) misleading, he might have been blameless in doing evil. It depends on whether he nonculpably believed that he was doing no wrong.

It is true that whether an evildoer is blameworthy depends in part on whether he believed nonculpably that he was doing no wrong. Perhaps the agent was not blameworthy for torturing the animals, if he believed nonculpably that he was doing no wrong. But to say that he did not even do anything morally wrong when he tortured the animals and enjoyed doing it seems to let him off the moral hook all too easily.

6. The Rationality Account Applied to Evaluative Uncertainty

If we reject Zimmerman's solution to cases of evaluative uncertainty, what shall we do? One option is to invoke rationality again and apply the notion of rational choice, given the agent's credence and preference, to cases of evaluative uncertainty as well (Bykvist 2014). The basic idea is that the conscientious agent cares about the value of outcomes in the sense that, for all value hypotheses V, he prefers the outcome x under hypothesis V to outcome y under the same hypothesis iff x is better than y, according to V; and she is indifferent between x under V and y under the same hypothesis iff x is equally as good as y, according to V. Furthermore, the strength of the agent's preference for x under V over y under V is proportionate to the value difference between x and y, according to V.

Here is how the rationality approach applies to Zimmerman's example about animal versus human well-being. The morally sensitive agent's preference for the outcome (speciesism, giving the medicine to John) over the outcome (speciesism, giving the medicine to Jane) is greater than his preference for (impartialism, giving the medicine to Jane) over (impartialism, giving the medicine to John). But then the rational choice, given the agent's preferences and beliefs, is to give the medicine to John, since, as a rational person he prefers bringing about the prospect (x, 0.5, y) rather than (z, 0.5, u) if his preference for x over z is much stronger than his preference for u over y. Assuming impartialism to be true, this would be a case in which the agent can't choose rationally without doing something that is morally wrong.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The obvious advantages of this rationality approach over Zimmerman's are the following. First, our evidence for evaluative hypothesis is relevant for rational choice, not for moral rightness, so we can take intrinsic value seriously and maintain that moral rightness depends on actual intrinsic value. Second, we cannot be morally obligated to do evil just because we have misleading evidence for some absurd evaluative hypothesis.

The rationality approach is also superior to the "just bite the bullet" approach, since it would not just say that we ought to do what in fact maximizes value, no matter what evidence you have about the values of outcomes. On this account, there is something further of normative significance we can say about cases of evaluative uncertainty, and that is that some actions are rational.

(p. 325) Whether it is superior to the "invoke different senses of 'ought'" approach depends on how this approach is spelled out. In fact, as in the case of empirical uncertainty, the rationality approach need not be seen as a competitor to this approach, since one can define a new nonobjective sense of "ought" in terms of rational action in a way analogous to what we did in the case of empirical uncertainty. For example, one can say that if we choose to give the medicine to Jane, then, no matter whether you care about outcome-values or not, you are taking an *evaluative* risk that a morally conscientious and rational person would not do in your situation, if she shared your credences about *evaluative* matters. Hence, it is to do something that an agent, who has a certain moral virtue, would not do in your situation. This account thus gives us a kind of nonobjective *moral ought*: what one ought to do in order to act virtuously in this respect, which is based on evaluative evidence rather than evaluative facts.

But there are problems too with the rationality account.

(a) Evaluative fetishism. I have described the morally conscientious person as caring (de dicto) about values of outcomes, for instance, as preferring a good outcome to a bad one. One could complain that this characterization makes him look like an evaluative fetishist.[8] Shouldn't the morally conscientious person care instead about what makes actions good or bad? Shouldn't he, for instance, care about human and animal well-being rather than just degrees of goodness?

In reply, one could point out that the objection presupposes a false dichotomy: you care either about goodness (badness) or about what makes outcomes good (bad). But it is possible and morally commendable to care about both. An agent who cares intrinsically only about goodness (badness) of outcomes seems deficient: she should also care about what makes outcomes good (bad), where this is read *de re*. But, equally, an agent who cares intrinsically only about the well-being of individuals and not at all about whether the outcomes of her actions are good or bad would also be deficient. The mark of a good moral agent is that she cares intrinsically not just about what makes the outcomes good (bad) but also about whether the outcomes are good (bad). Indeed, it is this fundamental concern for goodness and badness that explains why a good moral agent is expected to

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

change her motivation to act in the wake of a change in her fundamental evaluative principles.

How much should the agent care about the good-makers as compared to the goodness of the outcomes, if she is uncertain about what the good-makers are or how much goodness they make? Since we have assumed that she is morally conscientious and thus has a fundamental concern for goodness and badness, her intrinsic preference concerning prospects of possible good-makers should track her intrinsic preferences concerning prospects of value. This means that in the case of John and Jane, the agent should intrinsically prefer the well-being prospect of giving the pill to John over the well-being prospect (p. 326) of giving the pill to Jane, since she intrinsically prefers value prospect of giving the pill to John (100, 0.5, 100) to the value prospect of giving the pill to Jane (120, 0.5, 20).[9]

(b) *Intertheoretical comparisons of value*. I have assumed that it makes sense to compare the outcome-values across different evaluative hypotheses, but does it make sense? This is a pressing question since if these comparisons do not make sense, then we can no longer talk about the agent's preferences being sensitive to the value differences in cases of fundamental evaluative uncertainty. There are many proposals out there about how this can be done. To adjudicate between these theories, or to propose a different one, would require a paper of its own (Bykvist 2017; Bykvist 2020; Lockhardt 2000, 84–86; Ross 2006, 762–765; Sepielli 2009). Suffice it to say that Zimmerman is in exactly the same boat, since he also has to assume it makes sense. So his revised version of consequentialism is in no better position in this regard.

It should also be noted that for the rationality approach to have practical relevance I only need to assume that it is possible to compare value differences across *some* theories. I do not need to make the much stronger and controversial assumption that for *any* two theories, no matter how different, it makes sense to compare differences in value across these theories. The weaker assumption has some intuitive support. It is intuitive to say "if impartialism is true, completely curing the animal is somewhat better than partially curing the human, whereas, if speciesism is true, partially curing the human is considerably better." Comparative judgements like this one would all be mistaken if we could *never* compare value differences across theories.[10]

(c) *Rational uncertainty*. One could claim that by invoking the notion of rational choice, I have to face a new uncertainty problem: what is the rational thing to do when you don't know what it is rational to do? I have argued that in cases of uncertainty about empirical or moral matters what it is rational to (prefer to) do need not coincide with what it is morally right to do. One could object that this move will only provoke the (p. 327) further question about what to do in cases of uncertainty about rational matters.[11] This is a tricky question, which I do not have space to discuss here (but see Bykvist 2020), but note that anyone who believes in the notion of rational action has to confront it. It has nothing in particular to do with how we deal with evaluative uncertainty. Even if you reject my way of dealing with evaluative uncertainty, you still need to know what it is rational to prefer

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

given doubts about what makes a preference rational. After all, rational preference has a life to live outside evaluative uncertainty. For example, you can be uncertain about what rationality requires of you in cases of empirical uncertainty.

7. Uncertainty for Nonconsequentialists

As we have seen, the ignorance challenge is something the consequentialist must take seriously, but it is not clear what the best response is, for each approach has its own problems. Now this might be seen as something nonconsequentialists can use against consequentialism. This would be a mistake, however, since, as we shall see, adopting a nonconsequentialist theory does not take us out of the woods. Nonconsequentialist theories also need to grapple with uncertainty, both empirical and moral. Let us first list some considerations that show why they too have a problem with empirical uncertainty (Bykvist 2009a, 79–84).

(a) *The importance of future suffering.* Any reasonable moral theory must contain some principle of beneficence or nonmalevolence, saying that it is prima facie wrong to cause future suffering. But to know whether your action will cause future suffering, you might have to look into the distant future—some suffering might occur hundreds of years from now. Think again of the consequences of burying dangerous atomic waste in the ground. So, any reasonable moral theory that pays attention to future suffering will have to tackle the problem of knowing the future.

(b) *Knowledge about the past.* Many nonconsequentialist theories are backward looking. They tell you to keep past promises, to compensate for past wrongdoings, to give people what they deserve based on what happened to them in the past or what they did in the past. Therefore, knowledge about which action fulfils a duty requires knowledge about the past, but this might not be that easy to come by. For example, Robert Nozick defends an idea about distributive justice according to which a distribution is fair and ought to be maintained if it resulted from a chain of fair transactions, stretching into the past and ending at a point where someone justly appropriated some unowned material resources. Now, some of these transactions might have taken place way back in the past, (p. 328) and hence it might be virtually impossible for us now to know if all the transactions in the chain were fair.

(c) *Knowledge about your own psychological make-up.* For many nonconsequentialists, the psychological make-up of the agent matters a great deal for the moral rightness of his action. For instance, whether an effect is intended or only foreseen matters a great deal for rightness, according to the double-effect doctrine. You are permitted to do good, knowing and foreseeing that this will cause something bad, but you are not permitted to do something bad in order to do good. But this means that if I want to know whether my action is permissible, I need to know my intentions and beliefs, but that is not always easy. A doctor can, for instance, be confident that she is administering a painful drug to a

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

patient just because she wants to cure the patient, when in fact she does it partly because she takes pleasure in the mild suffering of an annoying patient.

Virtue theories will have similar problems, if they claim that the right action is the one that expresses a virtuous motive of the agent. The famous rock star who helps starving children might think he is acting on purely benevolent motives when in fact he is mainly after some social credit that will help his record sales. This lack of self-knowledge need not be easy to overcome. Some Kantian theories will of course be in a similar position, since to know whether an action is right, according to these theories, you need to know whether the fundamental maxim you are acting on could be rationally willed to become a universal law. But, as Kant himself acknowledged, it is not always transparent to us which maxim we are acting on.

(d) *Knowledge about external features of the action other than its effects.* Kantian theories will not just have problems with lack of self-knowledge. Even if you know which maxim you are acting on, you may not know whether your maxim could rationally be willed to become a universal law, for this requires very good imaginative and logical powers.

Nonconsequentialists not only have to address problems caused by ignorance of empirical matters. They also have to deal with uncertainty about fundamental *moral* matters. One of the prime examples is the problem of how to weigh prima facie duties. The agent can be ignorant about how to weight different prima facie duties. She can be uncertain about whether the prima facie duty not to break promises is more stringent than the prima facie duty not to cause minor physical harm.

Another example is the problem of following a virtue theory that judges an action right just in case it would be performed by a fully virtuous person. If you want to apply this theory, you will need to know what a fully virtuous person would do. But this knowledge is not easy to come by, for this requires moral knowledge about which character traits are virtuous and how they come together to form the unified character of a fully virtuous person.

What these examples show is that the ignorance problem is far from unique to consequentialism. Only very simple-minded and obviously false moral theories can easily be used in the sense that you can always find out what you ought to do and then successfully act on this knowledge. For instance, a theory that tells you to do what you feel like doing is easy to use but obviously false. Nonconsequentialists need to address this problem as well. They too can make use of any of the approaches that were listed for (p. 329) the consequentialist: just bite the bullet, invoke different senses of “ought,” change the ought-makers to something more subjective or epistemic, or invoke rationality and focus on what would be rational to do, given the agent’s credences and moral preferences (here a conscientious agent can be assumed to prefer rightdoings over wrongdoings, and minor wrongdoings over major wrongdoings).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

8. Concluding Remarks

The ignorance challenge is a serious challenge for consequentialism, and there are no obvious ways to deal with it. The "just bite the bullet" approach can be set aside, since it is obviously inadequate. The "changing the ought-maker" approach seems also very problematic, especially since the proposed new ought-makers do not seem to be the right ones. Whether we should go for the "invoke different senses of ought" approach or the "invoke rationality" approach is a more difficult matter, since there are so many different versions of the former approach. However, as I have argued, it is not clear that the two approaches must be seen as competitors. So we might not need to make a choice here.

These difficulties for consequentialism need to be put in perspective, however. The ignorance challenge is not a challenge only for consequentialism, even if the challenge for this theory takes a very stark form. Any reasonable moral theory will have to take into account the fact that ordinary agents are often in the dark about both normatively relevant empirical matters and fundamental moral matters.

References

Bykvist, K. 2009a. *Utilitarianism: A Guide for the Perplexed*. New York: Continuum.

Bykvist, K. 2009b. "Objective versus Subjective Moral Oughts." *Uppsala Philosophical Studies* 57.

Bykvist, K. 2011. "How To Do Wrong Knowingly and Get Away with It." *Uppsala Philosophical Studies* 58.

Bykvist, K. 2014. "Evaluative Uncertainty and Consequentialist Environmental Ethics." In *Environmental Ethics and Consequentialism*, edited by Leonard Kahn and Avram Hiller, 122–135. New York: Routledge.

Bykvist, K. 2017. "Moral Uncertainty." *Philosophy Compass* 12: e12408.

Bykvist, K. 2018. "Some Critical Comments on Michael Zimmerman's *Ignorance and Moral Obligation*." *Journal of Moral Philosophy* 15: 383–400.

Bykvist, K. 2020. *Moral Uncertainty*. With Toby Ord and Will MacAskill. Oxford: Oxford University Press.

Feldman, F. 2006. "Actual Utility, the Objection from Impracticality, and the Move to Expected Utility." *Philosophical Studies* 129: 49–79.

Graham, P. 2010. "In Defence of Objectivism about Moral Obligation." *Ethics* 121: 88–115.

Jackson, F. 1991. "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics* 101: 461–482.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(p. 330) Lockhardt, T. 2000. *Moral Uncertainty and Its Consequences*. Oxford: Oxford University Press.

Moore, G. E. 1912. *Ethics*. Oxford: Oxford University Press.

Oddie, G., and Menzies, P. 1992. "An Objectivist Guide to Subjective Value." *Ethics* 102, no. 3: 512–533.

Parfit, D. 2011. *On What Matters*. Oxford: Oxford University Press.

Portmore, D. 2011. *Commonsense Consequentialism*. Oxford: Oxford University Press.

Regan, D. 1980. *Utilitarianism and Co-operation*. Oxford: Clarendon Press.

Ross, J. 2006. "Rejecting Ethical Deflationism." *Ethics* 116: 742–768.

Ryle, G. 1949. *The Concept of Mind*. New York: Barnes and Noble.

Schroeder, M. 2007. *Slaves of Passions*. Oxford: Oxford University Press.

Sepielli, A. 2009. "What to Do When You Don't Know What to Do." In *Oxford Studies in Metaethics* 4: 5–28. Oxford: Oxford University Press.

Sepielli, Andrew. 2013. "What To Do When You Do Not Know What To Do When You Do Not Know What To Do…." *Noûs* 48, no. 3: 521–544.

Smith, M. 1996. *The Moral Problem*. Oxford: Blackwell.

Smith, M. 2006. "Moore on the Right, the Good, and Uncertainty." In *Metaethics after Moore*, edited by T. Horgan and M. Timmons, 133–148. Oxford: Oxford University Press.

Smith, M. 2009. "Consequentialism and the Nearest and Dearest Objection." In *Minds, Worlds, and Conditionals*, edited by I. Ravenscroft, 237–266. Oxford: Oxford University Press.

Zimmerman, M. 2008. *Living with Uncertainty*. Cambridge: Cambridge University Press.

Zimmerman, M. 2014. *Ignorance and Moral Obligation*. Oxford: Oxford University Press.

Notes:

([1]) For discussion about how to deal with both kinds of uncertainty simultaneously, see M. Smith, "Moore on the Right, the Good and Uncertainty," in *Metaethics after Moore*, edited by Horgan, T. and Timmons, M. (Oxford: Oxford University Press, 2006): 133–148; D. Portmore, *Commonsense Consequentialism* (Oxford: Oxford University Press, 2011), chap. 1; and M. Zimmerman, *Living with Uncertainty* (Cambridge: Cambridge University Press, 2008).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([2]) F. Jackson (1991). “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection,” *Ethics* 101 (1991): 461–482. I have relabeled the options. Examples of the same structure can be found in D. Regan, *Utilitarianism and Co-operation* (Oxford: Clarendon Press, 1980), 264–265; and in D. Parfit, *On What Matters* (Oxford: Oxford University Press, 2011), 159–160.

([3]) Jackson’s “Decision-theoretic Consequentialism” defends a similar principle.

([4]) The distinction between the two kinds of action-guidingness can be found in Zimmerman, *Ignorance and Moral Obligation*, who in turn follows the lead of G. Ryle, *The Concept of Mind* (New York: Barnes and Noble, 1949), 149–150.

([5]) The agent can also be uncertain about which rationality principle is correct. Should the agent hedge her bets at this level, too? If so, does that not lead to a regress? This is a serious issue that needs to be addressed. It should be noted, however, that this is *everyone’s* problem, not just those who take evaluative uncertainty seriously. After all, there can be rational uncertainty without evaluative uncertainty. For example, you can be uncertain about what rationality requires of you in cases of empirical uncertainty. For some suggestions on how to deal with this potential regress, see Andrew Sepielli, “What to do when you do not know what to do when you do not know what to do,” *Noûs* 48, no. 3 (2014): 521–544.

([6]) This fits the standard virtue ethical formula, which can be traced back to Aristotle: an action is virtuous (in one respect) just in case a virtuous person (with a certain virtue) would do it. Another instance, famous from Kant’s writings, is the egoist shopkeeper who always gives correct change because it promotes his own financial good and yet acts honestly in the sense that he does what a fully honest person would do in the circumstances.

([7]) This is not to deny that we can also make *internal* virtue assessments of both agents and acts. For example, an agent is kind only if she cares about other people. An act is caring only if the agent cares about others.

([8]) See M. Smith, *The Moral Problem* (Oxford: Blackwell, 1996), for a discussion of a deontic version of the fetishism problem.

([9]) One could take issue with this characterization of a morally conscientious agent’s preferences, because of what it entails for Huckleberry Finn–like cases: it seems like Huck should prefer the well-being prospect of turning Jim in, if he believes that Jim has less value because he is a slave. However, it is important to remember that to say that an agent is morally conscientious is to provide an *internal* assessment of the agent in terms of how well his moral beliefs and moral preferences hang together. We can also assess agent externally in terms of whether their preferences track the true values. On this score, Huck did very well. He also did the morally right thing (assuming that not turning him in maximized overall value).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([10]) Even in cases where we cannot make any intertheoretical comparisons of value, the rationality approach need not be silent. The agent's preferences can still be *constrained* by the value hypotheses she considers. More exactly, if a considered hypothesis V1 values outcome x over y, then the agent prefers x under V1 to y under V1; and if a considered hypothesis V1 values x and y equally, then the agent is indifferent between x under V1 and y under V1. The agent is free to form any preference (including indifference) between x under V1 and y under V2, if V1 is not identical to V2, on the condition that she does not thereby create an incoherent preference ordering (a violation of transitivity, for example). Rational action is defined relative to these preferences and the agent's credences for evaluative hypotheses.

([11]) And even if we can formulate principles for rational uncertainty, can't we be uncertain about them too? If so, we seem to have started on a regress. This is a serious issue that needs to be addressed. For some suggestions on how to deal with this potential regress, see Andrew Sepielli, "What to do when you do not know what to do when you do not know what to do," *Noûs* 48, no. 3 (2014): 521–544.

Krister Bykvist

Krister Bykvist is Professor of Practical Philosophy at Stockholm University, and Research Fellow at the Institute for Futures Studies, Stockholm. His primary interests are in moral philosophy broadly conceived. Most of his research is on topics in normative ethics, including consequentialism, utilitarianism, population ethics, climate ethics, prudence, and well-being. In metaethics, he has done work on noncognitivism, the nature of intrinsic goodness, and the normativity of mental states. More recently, he has done work on moral uncertainty. He has coauthored a book on this topic entitled Moral Uncertainty, to be published by Oxford University Press (release date February 2020). He has also written a book on utilitarianism, entitled Utilitarianism: A Guide for the Perplexed (Continuum, 2010).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Consequentialism and Indeterminacy

Caspar Hare

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.9

Abstract and Keywords

Many actions have indeterminate consequences. There is no fact of the matter about precisely what would have happened if they had not been taken. How should act consequentialists think about their objective moral status? This paper suggests a way for act consequentialists to think, and it draws attention to three remaining problems for act consequentialists to solve.

Keywords: conditionals, conditional indeterminancy, consequentialism, outcomes.

1. Outcomes

ACT consequentialists say that the objective moral status of any action is determined by whether its outcome is relevantly better or worse than the outcomes of alternatives to it. That's well and good for a rough sketch of a moral theory, but there are many details to fill in. What are *actions*? What is *objective moral status*? For any given action, what is its *outcome*? What is it for one outcome to be *relevantly better or worse* than another? For any given action, what are the *alternatives to it*? How, exactly, is the objective moral status of an action *determined* by these things?

I will focus on one little bit of the detailed picture here. For any given action, what is its *outcome*?

Back in the day some people working in the tradition then known as "utilitarian," now known as "consequentialist," may have thought of the outcome of an action as including all and only things caused by the action.[1] So yesterday afternoon, the afternoon of July 22, 2019, I scratched my nose. Subsequent to that, many things happened. I felt a pulse of sweet relief. Boris Johnson became Prime Minister of the United Kingdom. The pulse of relief was part of the outcome of my action, because it was caused by my action. But Johnson's ascension was not, because it wasn't caused by my action.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Nowadays consequentialists tend to think of the outcome of an action as including all and only the things that would happen if the action were taken. On this way of thinking the outcome of my scratching my nose includes my pulse of relief and Johnson's ascension, along with everything that has happened, is happening, or will happen: the Big (p. 345) Bang, the evolution of the opposable thumb, the French Revolution, the evaporation of the waters of the Earth, Mercury's being swallowed up by the Sun, … *everything*.

There are at least three reasons to prefer the modern approach. First, from the point of view of theoretical economy, it enables us to avoid an unwieldy sort of double comparison. On the old-fashioned approach, to establish what the outcome of an action is, I must establish what it caused to happen, which effectively requires me to compare the action to its alternatives (with respect to what would happen if they were taken). Then, to establish what the moral status of an action is, I must compare the action to its alternatives a second time (with respect to whether it or its alternatives has a relevantly better outcome). On the modern approach there's just one comparison to be made: outcome to outcomes.

Second, the modern approach enables act consequentialism to accommodate more complex evaluative judgments without stretching the notion of causal consequence beyond its natural limit. An act consequentialist may want to say, for example, that sometimes whether you ought to punish someone depends on whether he committed a crime. Other things being equal, it is bad to punish someone for a crime he did not commit, but good to punish someone for a crime he committed. So sometimes events uncaused by an action may have a bearing on its moral status. To accommodate this on the old-fashioned approach, we must say that one of the things you cause, when you punish the innocent person, is *somebody being punished for a crime he did not commit*. Facts about events you do not cause (in this case *the person not having committed the crime*) are, so to speak, baked in to events that you cause. But this is inelegant.

Third, the modern approach more cleanly captures the motivating idea behind act consequentialism. What fundamentally matters is *all that ever happens*. So the important question with respect to an action is this: "what is all that would ever happen if this action were taken?"

Well and good, but there's a problem. What if, sometimes, for some actions, there is no fact of the matter about precisely what would happen if they were taken? What if *conditional indeterminacy* is a real thing?

2. Conditional Indeterminacy

Sometimes, when we ask what would happen (or would have happened) if a certain condition obtains (or has obtained), we get a curious pattern of truths. A proposition of this form is true:

> If A, then it would be that B or C.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

But neither proposition of these forms is true:

> If A, then it would be that B.
>
> If A, then it would be that C.

(p. 346) One situation in which this pattern can arise involves indeterministic laws. Consider:

> *The Truncated Quantum Mechanics Lecture*
>
> I am teaching a class on quantum mechanics. As a final demonstration to the class, I plan to fire a single photon through a narrow slit and measure its trajectory. The laws of quantum mechanics determine that it will angle up or angle down, but they do not determine which. Sadly I run out of time. I never perform my demonstration.

Now this is true:

> If I had fired the photon, then it would have angled up or angled down.

But neither of these is true:

> If I had fired the photon, then it would have angled up.
>
> If I had fired the photon, then it would have angled down.

Why? Well, the short version of the answer is this: Modal propositions (those that have to do with what might be the case, could be the case, or would be the case) are made true or false by nonmodal propositions (those that have to do with what is, has, or will be the case). But there are no nonmodal propositions that make one, but not the other, of this pair true. It is consistent with the way our world is, and the laws that govern our world, that the photon would have angled up, and it is consistent with the way our world is, and the laws that govern our world, that the photon would have angled down.

The longer version of the answer is this: On the now standard approach to evaluating the truth of conditionals,[2] a proposition of the form

> If it had been that A, then it would have been that B.

is true just in case some possible world (where possible worlds represent fully determinate ways for the world to be, nonmodally speaking) in which A and B are both true is closer (where distance is a measure of relevant similarity) to the actual world than any possible worlds in which A is true but B is not. Applying the standard approach to the three conditionals earlier, some world in which I fire the photon and it either angles up or angles down is closer to the actual world than all worlds in which I fire the photon and it does not angle up or down, so the first conditional comes out true. But no world in which I fire the photon and it angles up is closer to the actual worlds than all worlds in which I

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

fire the photon and it angles down, and vice-versa, so neither of the second and third conditionals comes out true.[3]

(p. 347) In situations like this I will say that that there is *no fact of the matter* about whether the situation described by the consequent (e.g., the photon angling up) would obtain. And I will refer to the general phenomenon as *conditional indeterminacy*.

Conditional indeterminacy need not arise from indeterminacy in physical laws. Consider:

> *The Balance of Justice*
>
> You stand before a scale-of-justice-style balancing scale, with two equal weights nicely balanced on either side.

Now this is true:

> If one of the weights were substantially heavier than the other, then the scale would tip left or tip right.

But neither of these is true:

> If one of the weights were substantially heavier than the other, then the scale would tip left.
>
> If one of the weights were substantially heavier than the other, then the scale would tip right.

No nonmodal propositions make one, but not the other, of the second and third modal propositions true. No possible world in which the left weight is substantially heavier, and the scale tips left, is relevantly more similar to the actual world than all worlds in which the right weight is substantially heavier, and the scale tips right. This is not because of indeterministic physical laws. It is because the common antecedent of the second and third conditionals, "If one of the weights were substantially heavier than the other … ," is underspecified in relation to their consequents. If the left weight were heavier, then which way would the scale tip? There's a good answer to that question—it would tip left. If the right weight were heavier, then which way would the scale tip? There's a good answer to that question—it would tip right. If one of the weights were heavier, then which way would the scale tip? There's no good answer to that question.

Can conditional indeterminacy due to conditional underspecification arise when the antecedent of the conditional describes a human action? It would appear so. Consider

> *The Sensitive, Accurate Kitchen Scale*
>
> There's a kitchen-style electronic scale before me. It gives an accurate reading in newtons, to four decimal places. I mull over whether to press my thumb down on the scale, but decide not to. I leave the scale entirely alone.

Now this is true:

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> If I had pressed my thumb down on the scale, then the final digit of its maximum reading would have been even or odd.

(p. 348) But neither of these is true:

> If I had pressed my thumb down on the scale, then the final digit of its maximum reading would have been even.
>
> If I had pressed my thumb down on the scale, then the final digit of its maximum reading would have been odd.

There are many nearby possible worlds in which I press down on the scale and get different maximum readings. In some, the final digit is even; in others, odd. But no world in which I press and get an even final digit is relevantly more similar to the actual world than all worlds in which I press and get an odd final digit, and vice versa. The antecedent of the pair of conditionals above, "If I had pressed my thumb down on the scale …," is underspecified relative to their consequents. If I had pressed down with exactly 3.4068 N of force, would the final digit have been even or odd? There's a good answer to that question. It would have been even. If I had pressed down with exactly 3.4067 N of force, would the final digit have been even or odd? There's a good answer to that question. It would have been odd. If I had pressed down, would the final digit have been even or odd? There's no good answer to that question. Whether the final digit would have been even or odd depends on precisely how I would have pressed down, if I had pressed down. But there is no fact of the matter about precisely how I would have pressed down, if I had pressed down.

How widespread is this phenomenon? Two further examples should illustrate that it is very widespread indeed. Consider:

> *The Unflipped Coin*
>
> You balance a coin on your thumb, think about whether to flip it, decide against doing so, and return it to your pocket.

Now this is true:

> If you had flipped the coin, then it would have landed heads or landed tails.

But neither of these is true:

> If you had flipped the coin, then it would have landed heads.
>
> If you had flipped the coin, then it would have landed tails.

There are many nearby possible worlds in which you flip the coin at many different heights, many different angular velocities, many different rates of rotation. In some you get heads; in others you get tails. But no world in which you flip the coin and get a head is relevantly more similar to the actual world than all worlds in which you flip the coin

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

and get a tail, and vice-versa. The antecedent of the aforementioned conditionals, "If you had flipped the coin … ," is underspecified relative to their consequents. Whether or not you would have gotten a head or tail depends on precisely how you would have flipped the (p. 349) coin, if you had flipped the coin. But there is no fact of the matter about precisely how you would have flipped the coin, if you had flipped the coin.

Finally consider:

> *The Unwaved Hand*
>
> Shall I wave at the marching band as they process by me? I don't do it. And exactly 100 days later it rains in Cambridge, MA.

Now this is true:

> If I had waved my hand, then 100 days later it would have rained or not rained in Cambridge, MA.

But chaotic behavior in our best models of the atmosphere suggests that neither of these is true:[4]

> If I had waved my hand, then 100 days later it would have rained in Cambridge, MA.
>
> If I had waved my hand, then 100 days later it would not have rained in Cambridge, MA.

The antecedent of these conditionals, "If I had waved my hand …" is underspecified relative to their consequents. Whether or not it would have rained in Cambridge, MA, depends on precisely how I would have waved my hand, if I had waved my hand. But there is no fact of the matter about precisely how I would have waved my hand, if I had waved my hand.

3. Accommodating Conditional Indeterminacy

What can an act consequentialist do about conditional indeterminacy? One approach might be to allow the indeterminacy to infect the objective moral status of actions. When there is no fact of the matter about would have happened if an action had been taken, then it is indeterminate what the outcome of the action is. When it is indeterminate what the outcome of the action is, it may be indeterminate whether its outcome is in the relevant way better than the outcomes of alternatives to it, and hence indeterminate what its moral status is. That's just how it goes, sometimes. So, for example, in the Unflipped Coin case, it is indeterminate whether the outcome of your flipping the coin is one in which the coin lands heads or one in which the coin lands tails. And supposing that it would be relevantly *wonderful* if you were to flip the coin and it land heads, and (p. 350) relevantly *terri-*

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ble if you were to flip the coin and it land tails, then it is indeterminate what the objective moral status of your flipping the coin is.

But this is not so satisfactory. It suggests that if chaos theorists are right about the weather, then pretty much all actions have indeterminate objective moral status. For pretty much anything I might do now, in 2019, it is indeterminate whether the outcome of that action includes tornados sweeping through the city of Tulsa in September 2020. So (on the reasonable assumption that, other things being equal, outcomes in which tornados sweep through Tulsa are relevantly much worse than outcomes that don't) it is indeterminate what the objective moral status of the action is. But we moral theorists hope to say something more discriminate than that "everything we ever do is indeterminately right, indeterminately wrong, indeterminately such that it ought to be done, indeterminately such that it ought not to be done." If we are forced to say that, then so be it, but let's see if there is something else to say.

Another approach might be to let the outcome of an action be a collection of fully specific world histories (think of these things as Lewisian possible worlds, if you are familiar with such things), each one representing a way things might go if the action were to be taken. So, given that it is determinate that the coin would travel through the air, if flipped, but indeterminate whether it would land heads or tails, if flipped, all worlds in the outcome of flipping the coin are worlds in which the coin travels through the air, some worlds in the outcome are worlds in which it lands heads, some worlds in the outcome are worlds in which it lands tails.

On a crude implementation of this approach we leave it at that. On a more sophisticated implementation, we go on to associate probabilities with classes of possibilities in the outcome. So, in the Sensitive, Accurate Kitchen Scale case, though there is no fact of the matter about whether the final digit of the scale's reading would have been even or odd, if I had pressed it, the class of worlds in which I press and the final digit comes out even has the number 0.5 associated with it—representing the fact that the conditional probability of the final digit being even, were I to press it, is 0.5.

With these conditional probabilities in place, we can use the formidable resources of expected value theory to assign values to outcomes. First, we divide the outcome into classes of similarly valued possibilities. Next, we assign a number to each class, representing the value of each of the possibilities in it. We weight that number by the conditional probability associated with the class. We sum. We get the value of the outcome.

Are act consequentialists home free? Not quite. In what remains of this chapter I will describe three remaining problems for the act consequentialist. I have ideas about how to solve the first problem. The second and third problems I will leave to you.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

4. What If Causation Matters?

The first problem has to do with incorporating conditional indeterminacy into a ranking of possibilities. Some act consequentialists may want to say that causal facts can have (p. 351) a bearing on whether one possibility is relevantly better than another. Sometimes, in order to know which outcome is relevantly better, we need to need to know more than what happens in each; we need to know what causes what to happen in each. They may want to say, for example, that it is relevantly better, other things being equal, that human success be caused by human effort. It is relevantly better that you study hard and pass your exam as a result of your study than that you study hard and pass your exam by chance.[5] They may want to say, for another example, that it is relevantly worse, other things being equal, that human misery be caused by malicious human action. It is relevantly worse that you be maimed by a boulder dislodged from on high by an enemy than that you be maimed by a boulder dislodged from on high by wind and hail.

Given that causation is very important to this sort of act consequentialist (call her a *sensitive-to-causes consequentialist*, or *STC consequentialist*, for short), she now owes us an account of how causation works when conditionals are indeterminate. Consider, for example:

> *The Stolen Coin*
>
> For reasons too tedious to itemize here (embellish the story as you like so as to make this all true), nature has conspired to put me in a terrible situation: I am heading precipitously toward death. My only hope of salvation lies with a coin in my pocket. If I flip it and get a head, then I will live. Otherwise I will die. But my enemy takes the coin from me before I get a chance to flip it. I die.

The STC consequentialist wants to say that my death, relevantly bad in any case, is relevantly worse if my enemy caused it to happen by taking the coin. But did my enemy cause my death to happen by taking the coin? I died. There is no fact of the matter about whether I would have lived or died if my enemy had not taken the coin. The conditional probability of my dying if my enemy were not to have taken the coin is ½. Does all that make for causation?

The STC consequentialist might want look to the philosophical literature on causation for help with this question.

On David Lewis's theory of causation in what he called a "chancy world," we cause things to happen by significantly raising the chances of their happening—with what counts as "significant" varying from context to context.[6] In this case, after my enemy stole the coin I was certain to die, but if my enemy had not stolen the coin, then I would have had a ½ chance of dying. By stealing the coin, my enemy doubled the chances of my dying. Is that significant? It is up to the sensitive-to-causes consequentialist who adopts Lewis's theory

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

to tell us whether it is. If it is, then mine is the worse sort of death, the sort of death caused by human action.

Johan Frick has recently criticized Lewis's theory on structural grounds (e.g., sometimes, on Lewis's account, two people cause something to happen without either person (p. 352) causing the something to happen—Frick says that cannot be).[7] Frick proposes instead that we cause things to happen by to any degree raising the chances of their happening. In this case you most definitely did to some degree raise the chances of my dying by stealing the coin, so you caused me to die. So an STC consequentialist who adopts Frick's theory will again say that this is the worse sort of death, the sort of death caused by human action.

But STC consequentialists should not adopt either of Lewis or Frick's theories. Consider:

> *Two Deaths, One Stolen Coin*
>
> Nature has conspired to put you and me in a terrible situation: We are heading precipitously toward death. Our only hope of salvation lies with a coin in my pocket. If I flip it and get a head, then I will live and you will die. If I flip it and get a tail, then you will live and I will die. Otherwise we both will die. But my enemy takes the coin from me before I get a chance to flip it. We both die.
>
> *Two Deaths, One Stolen Envelope*
>
> Again you and I are heading toward death. Our hope this time around lies with a sealed envelope. One of our names is written inside. If I open the envelope, then the named person will live and the unnamed person will die. Otherwise we both will die. But my enemy takes the envelope from me before I get a chance to open it. We both die. Though we never know it, as a matter of fact my name was written inside the envelope.

On Lewis's and Frick's theories, in Two Deaths, One Stolen Coin, my enemy causes two deaths, while in Two Deaths, One Stolen Envelope, my enemy causes just one death. So, given that other things are relevantly equal, STC consequentialists who adopt Lewis or Frick's theory must say that the outcome of Two Deaths, One Stolen Coin is worse than the outcome of Two Deaths, One Stolen Envelope. But they should not say this. They would thereby be saying that it matters whether conditionals are determinate or indeterminate. But it doesn't matter in this way whether conditionals are determinate or indeterminate.

If this isn't obvious to you, imagine yourself facing an awful choice.

> *Stealing an Envelope versus Stealing a Coin*
>
> You can prevent the apocalypse in either of two ways: You can steal a coin that will otherwise save one of two unknown-to-you-people (in the manner earlier). You can

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> steal an envelope that will otherwise save one of two different unknown-to-you people (in the manner earlier).

An STC consequentialist who adopts Lewis or Frick's theory must say that you ought to steal the envelope. If you steal the coin, then you cause the deaths of two people. If you steal the envelope, then you cause the death of one person. It is worse, other things being (p. 353) equal, that you cause two deaths than that you cause one death. But I submit to you that that's not right. It's fine for you to go either way here.

So what theory of chancy causation should an STC consequentialist adopt? That's our first problem.

Here's a proposal: She should say that the important sort of causal relations come in degrees. She should say that in Two Deaths, One Stolen Coin, my enemy causes my death to degree 0.5, while in Two Deaths, One Stolen Envelope, my enemy causes my death to degree 1. It is bad that people cause deaths to any degree. The badness varies by degree.

How exactly does the badness of causing to a degree aggregate? Other things being equal (in particular, the number of deaths being the same), is it better or worse, for example, that three deaths be caused to degree 0.25 than that one death be caused to degree 0.8? Questions about value aggregation are notoriously difficult to answer, but we have a guide to how to answer this one. If the outcome of Two Deaths, One Stolen Coin is exactly as bad as the outcome of Two Deaths, One Stolen Envelope, then it must be exactly as bad that two deaths be caused to degree 0.5 as that one death be caused to degree 1. More generally, if it does not matter whether conditionals are determinate or indeterminate, then it must be worse, other things being equal, that m deaths be caused to degree n than that i deaths be caused to degree k, just in case mn is greater than ik.

5. What Makes Some Worlds Closer Than Others?

The second problem has to do with relevant closeness. Sometimes it is not obvious how to assess conditionals. Consider, for example:

> If kangaroos had no tails, then they would topple over.[8]

> If kangaroos had no tails, then they would not topple over.

On one way of thinking the first is true. If all the kangaroos in the world were to lose their tails, through some terrible, coordinated mishap, then they would become unbalanced and topple forward. On another way of thinking the second is true. If kangaroos had not evolved powerful, weighty tails, then they would have evolved different forelegs and remained nicely balanced. On yet another way of thinking neither is true. This is a case of conditional indeterminacy due to conditional underspecification.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Which way of thinking is right? On the standard theory that depends on whether worlds in which kangaroos lose their tails and topple forward (call these the *topple-worlds*) are more relevantly similar to the actual world than worlds in which kangaroos (p. 354) evolve no tails and remain nicely balanced (call these the *no-topple worlds*). That depends on what "relevant similarity" is. And that changes from conversational context to conversational context. When you are among mechanical engineers, talking about mechanical engineering, you are right to say that tail-less kangaroos would topple over. In that context the topple-worlds are relevantly more similar to the actual world than the no-topple worlds. When you are among evolutionary biologists, talking about evolutionary biology, you are right to say that tail-less kangaroos would be nicely balanced. In that context the no-topple worlds are relevantly more similar to the actual world than the topple worlds. When you are with your friends, talking about nothing in particular, you are wrong to say that tail-less kangaroos would topple over, wrong to say that tail-less kangaroos would be nicely balanced. In that context neither topple nor no-topple worlds are relevantly more similar to the actual world.

Consider, for another example, a pair of conditionals having to do with the men's long jump at the Rio Olympics (actually won by Jeff Henderson, with a jump of 8.38 meters):

> If I had entered the men's long jump at the Rio Olympics and jumped between 8.20 and 8.56 meters, then I would have won gold.
>
> If I had entered the men's long jump at the Rio Olympics and jumped between 8.20 and 8.56 meters, then I would not have won gold.

On one way of thinking, neither conditional is true. Their antecedents are underspecified. On another way of thinking, the second is true. Wildly unrealistic at it may be, my jumping and just losing is a little closer to reality than my jumping and just winning.

Which way of thinking is right? On the standard theory, that depends on whether distant worlds in which I jump between 8.20 and 8.31 meters are, other things equal, relevantly just a bit closer to the actual world than distant worlds in which I jump between 8.32 and 8.50 meters. That depends on what relevant closeness is, and that may change from conversational context to conversational context.

The general phenomenon is this: Different groups of people are considering whether or not something, call it E, would happen if something else were to happen. When talking among themselves, one group says that E would happen, another group says that E would not happen, and yet another group says that there is no fact of the matter about whether E would happen. They are all right, because they are correctly applying different standards of evaluation to the counterfactuals, using different, equally legitimate measures of similarity.

In light of this phenomenon, we might want to ask: When we do moral philosophy, is there one right way to evaluate conditionals whose antecedent is a human action? The act consequentialist is in a bind here. She must either say there is or say there is not one similar-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ity relation that plays a special role in determining the moral status of an action. If there is not one special similarity relation, then she must say that actions are merely objectively right or wrong relative to similarity relations. If there is a special similarity relation, then she owes us an account of what it is. What is it? I know of no very successful efforts to address this question.

(p. 355) 6. What If There Are No Conditional Probabilities?

The third problem has to do with conditional probabilities. The act consequentialist treatment of conditional indeterminacy that I sketched in section 3 relied on the fact that sometimes, when events might have occurred, and there's no fact of the matter about whether they would have occurred, we can associate conditional probabilities with their occurring. How exactly does this work? The idea is that, when no world in which A and B is relevantly closer than all worlds in which A and not-B, and vice versa, we may nonetheless be able to say that a certain percentage of suitably nearby A-worlds are B worlds. Then we can say that the conditional probability of B, if it were that A, is that percentage. For example, in the Sensitive, Accurate Kitchen Scale case, though no world in which I press and get an even final digit is closer than all worlds in which I press and get an odd final digit, we can say that 50 percent of suitably nearby worlds in which I press are worlds in which I get an even final digit. So the conditional probability of my getting an even final digit, if I were to press, is 0.5.

But there are infinitely many suitably nearby worlds in which I press! How can it be that 50 percent of them are worlds in which I press and get an even final digit? What's 50 percent of infinity? The idea, roughly, is that though the set of suitably nearby worlds is infinite, it can be partitioned into finitely many cells. Some finite partitions are more *natural* than others—for example, the partition that divides the set into two cells, one containing those worlds in which I press down with more than 4 Newtons of force, and the other containing those worlds in which I press down with less than 4 Newtons of force, is more natural than the partition that divides the set into two cells, one containing those worlds in which I press down with between 4 and 4.1 Newtons of force, and the other containing those worlds in which I don't. And some finite partitions are more *fine-grained* than others—for example, a partition that divides the set into 100 cells is more fine-grained than a partition that divides the set into 2 cells. For any sufficiently natural, sufficiently fine-grained finite partition of the set of nearby worlds in which I press, 50 percent of the cells contain only worlds in which I get an even final digit. That is what we mean by saying that 50 percent of the cells in the infinite set are worlds in which I get an even final digit.

Well and good. But sometimes, for some instances of conditional indeterminacy, it just isn't possible to associate conditional probabilities with events occurring. Consider:

The Absent Cube

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> No cube of gold ever enters my room.

Now this is true:

> If a cube of gold, of side length between 0 and 2 meters, had materialized in my room (violating actual laws of physics), then its mass would have been above or below 19,300 kilograms.

(p. 356) And, due to conditional underspecification, neither of these is true:

> (Above) If a cube of gold, of side-length between 0 and 2 meters, had materialized in my room, then its mass would have been above 19,300 kilograms.
>
> (Below) If a cube of gold, of side-length between 0 and 2 meters, had materialized in my room, then its mass would have been below 19,300 kilograms.

A cubic meter of gold has a mass of 19,300 kilograms. So whether the cube would have had mass above or below 19,300 kilograms depends on whether it would have had a side-length above or below 1 meter. But there is no fact of the matter about whether it would have had side-length above or below 1 meter, if it had had side length between 0 and 2 meters.

But this time we cannot easily associate conditional probabilities with the pair. Here are three simple finite partitions of the set of suitably nearby worlds in which their antecedent holds:

> P1 A partition that divides the set into worlds in which the cube has side-length between 0 and 1 meters, and worlds in which the cube has side-length between 1 and 2 meters.
>
> P2 A partition that divides the set into worlds in which the cube has face-area between 0 and 2 square meters, and worlds in which the cube has face-area between 2 and 4 square meters.
>
> P3 A partition that divides the set into worlds in which the cube has volume between 0 and 4 cubic meters, and worlds in which the cube has volume between 4 and 8 cubic meters.

If partitions like P1 are more natural than partitions like P2 and partitions like P3, then the conditional probability of (Above) is 1/2. If partitions like P2 are most natural, then the conditional probability of (Above) is 3/4. If partitions like P3 are most natural, then the conditional probability of (Above) is 7/8. But really none of these partitions are more or less natural. So (Above) has no conditional probability.

> Call a conditional like (Above), with no conditional probability, *deeply indeterminate*.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Can conditionals whose antecedents are actions available to an agent, and whose consequents are matters of relevant significance, be deeply indeterminate? If so, then, when applied to these actions and their alternatives, act consequentialism looks to be in deep trouble. I see no good act consequentialist account of the objective moral status of such actions. The most an act consequentialist can say is: “If you were to do this, then a bunch of important things might or might not happen, with no probability. If you were to do that, then a bunch of important things might or might not happen, with no probability.” We must either bite the bullet, and accept that this class of actions has no objective moral status, or reject act consequentialism.

(p. 357) So the pressing problem, for an act consequentialist, is to tell us when conditionals are and are not deeply indeterminate, and to show that conditionals whose antecedents are the actions whose moral status we are interested in are not deeply indeterminate. Again, I know of no very successful efforts to address this problem.

7. Moving Forward

To summarize: Conditional indeterminacy is real. Once they have acknowledged this, act consequentialists have work to do.

You may want to say, “This isn’t my work. I am not an act consequentialist. I don’t think that the objective moral status of any action is determined by whether its outcome is relevantly better or worse than the outcome of alternatives to it.” But be careful. So long as you think that consequences matter at least a bit, so long as you think that at least sometimes the moral status of an action is at least in part determined by whether its outcome is relevantly better or worse than the outcome of alternatives to it, this is your work, too.

References

Carlson, E. 1995. *Consequentialism Reconsidered*. Dordrecht: Kluwer.

Frick, Johann. Unpublished manuscript. “Probabilistic Causation, Moral Responsibility, and the Problem of Aggregate Effects.”

Hare, Caspar. 2011. “Obligation and Regret When There Is No Fact of the Matter about What Would Have Happened If You Had Not Done What You Did.” *Noûs* 45: 190–206.

Hare, Caspar. Forthcoming. *Living in a Strange World*. Oxford University Press.

Lewis, David. 1973. *Counterfactuals*. Oxford: Blackwell.

Lewis, David. 1986. “Postscripts to ‘Causation.’” In his *Philosophical Papers: Volume II.* Oxford: Oxford University Press.

Moore, G. E. 1912. *Ethics*. New York: Henry Holt and Company.

Stalnaker, Robert. 1984. *Inquiry*. Cambridge, MA: MIT Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Velleman, David. 1991. “Well-Being and Time.” *Pacific Philosophical Quarterly* 72, no. 1: 48–77.

Notes:

([1]) This is what Eric Carlson (in Carlson 1995) calls the “Principle of Causal Outcomes.” Carlson attributes the principle to G. E. Moore in Moore (1912), though Carlson acknowledges that the strength of Moore’s commitment to the principle is somewhat unclear.

([2]) Canonically sourced to David Lewis and Robert Stalnaker in Lewis (1973) and Stalnaker (1984).

([3]) I should note that on Lewis’s version of the approach the second and third conditionals are false, while on Stalnaker’s version of the approach they each have indeterminate truth value. I prefer Lewis’s version, for reasons I explain in Hare (2011) and Hare (forthcoming), but this won’t matter as we move on.

([4]) See Hare (forthcoming).

([5]) This is explicit, for example, in David Velleman’s theory of the value of a life. See Velleman (1991).

([6]) See Lewis (1986).

([7]) See Frick (unpublished manuscript).

([8]) This is the first example of a conditional that Lewis gives in Lewis (1973). He does not use it to illustrate the phenomenon of underspecification, but it serves that purpose well.

Caspar Hare

Caspar Hare is Professor of Philosophy at the Massachusetts Institute of Technology. He writes about ethics, practical rationality, metaphysics, and about the connections between them. He is the author of two books: On Myself, and Other, Less Important Subjects (Princeton University Press, 2009) and The Limits of Kindness (Oxford University Press, 2013).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Deontic Pluralism and the Right Amount of Good

Richard Yetter Chappell

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.10

Abstract and Keywords

Consequentialist views have traditionally taken a *maximizing* form, requiring agents to bring about the very best outcome that they can. But this maximizing function may be questioned. *Satisficing* views instead allow agents to bring about any outcome that exceeds a satisfactory threshold or qualifies as "good enough." *Scalar* consequentialism, by contrast, eschews moral requirements altogether, instead evaluating acts in purely comparative terms, that is, as better or worse than their alternatives. After surveying the main considerations for and against each of these three views, I argue that the core insights of each are not (despite appearances) in conflict. Consequentialists should be *deontic pluralists* and accept a maximizing account of the *ought of most reason*, a satisficing account of *obligation*, and a scalar account of the *weight of reasons*.

Keywords: maximizing, satisficing, scalar consequentialism, obligation, ought, rightness, blameworthiness, deontic pluralism

1. Introduction

CONSEQUENTIALISM directs us to promote the good. But how much? Is there a "right amount" of good to produce—a level we need to reach in order to qualify as acting rightly? Maximizers hold the right amount to be the maximum available to the agent in the circumstances: you act wrongly if an alternative option would have brought about a better outcome. Satisficers identify a less strict threshold, allowing that some suboptimal acts may nonetheless be "good enough." Finally, scalar consequentialists reject the question, simply affirming that acts are better the more good that they produce.

In this paper, I'll question an assumption that underlies this whole debate: that there is a single sense of "rightness" about which these various forms of consequentialism disagree. Section 2 discusses maximizing consequentialism, with particular attention to the demandingness objection and to broader structural concerns with identifying rightness and optimality. Section 3 explores the case for scalar consequentialism, but then suggests two senses of "rightness" that the scalar theorist lacks good grounds for dismissing. Sec-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

tion 4 makes the case for satisficing consequentialism, showing how the view can be defended against three important objections. Finally, section 5 explains how *deontic pluralism* enables us to reconcile these three forms of consequentialism. We can accept an attractive package view that is scalar at core, maximizing about the ought of most reason and satisficing about obligation.

(p. 499) 2. Maximizing

Maximizing act consequentialists hold that an act is right if and only if it produces at least as much value as any other act that the agent could perform at that time. A maximizing approach to ethics can be motivated by appeal to our inclination toward maximizing accounts of practical rationality more generally (Scheffler 1985). Once appropriate practical goals have been identified, it would seem instrumentally rational to act so as to best achieve the relevant goals.[1]

There's no question that from a consequentialist perspective, a suboptimal act is *worse* than an optimal alternative. So, inasmuch as asking after "*the* right act" in a situation builds in a linguistic presumption of uniqueness—that there is just one choice among the agent's options which is the right one—a maximizing account of right action can come to seem very natural to a consequentialist.

Nonetheless, there are reasons to be wary of insisting that consequentialism take a maximizing form. Despite sometimes speaking of "the right act," we don't generally think that morality is so restrictive as to rule out almost every option we have in any given situation. We expect morality to rule out morally unacceptable options and leave us free to choose among the remainder—including, typically, many options that are considered acceptable despite falling short of moral perfection. Maximizers, by contrast, are committed to denying any space for such *moral autonomy* (though they can certainly defend the social practice of autonomy, given that trying to *force* people to act maximally well would plausibly backfire). We may choose how to break ties between morally optimal options, if there happens to be more than one, but we are given no further permissible options beyond that.

By so restricting our practically permissible options, maximizing consequentialism also proves conceptually restricting. For, in fleshing out the structure of our moral options, common sense recognizes conceptual room for the *supererogatory*, that is, acts that go "above and beyond" the call of duty. This possibility is entirely precluded by a maximizing account of our duties. If required to always do the best, there is no room left for us to do better than the minimum required.

Further, by requiring actions that are intuitively supererogatory, maximizers are subject to the objection that their conception of morality is *overly demanding*, or places unreasonable demands on agents. This objection is standardly developed in relation to the demands of beneficence, and especially Singer (1972)'s view that we should give to (p. 500)

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

the point of marginal utility—where any further donation would hurt us more than it would help the recipients.

How weighty should consequentialists find the demandingness objection? Norcross (2020, chap. 2.2) questions the reliability of anti-demandingness intuitions on the grounds of their being obviously self-serving for the culturally influential. Sobel (2007) counters it differently, arguing that the demandingness objection *presupposes*, rather than supports, nonconsequentialist asymmetries between (e.g.,) action and inaction. After all, however demanding maximizing consequentialism may seem for the affluent, we may wonder why alternative theories are not seen as even *more* demanding for the poor, given the much greater costs that will befall them if the affluent fail to give much aid (Murphy 2000, 55). Ordinary demandingness intuitions are evidently not tracking a neutral evaluation of the (net) costs to people of general compliance with a theory. This may support Norcross's suspicions and lead us to doubt whether we have principled reasons to care about whatever it is that they *are* tracking.

Defenders of the demandingness objection respond by stressing the significance of *self-imposed* costs as raising distinctive moral questions. We may ask whether agents have sufficient reason to comply with a putative moral demand that would prove very costly to them (Woollard 2016), or whether they would warrant blame should they fail to so comply (McElwee 2017). Such questions aren't applicable to moral patients when they suffer costs imposed from without. Morality does not, in this instance, demand that the sufferers *do* anything. And so the distinctive questions of moral demandingness simply don't arise in relation to such costs.

This account allows us to make sense of the special role that self-imposed costs play in our assessments of demandingness. But it remains an open question whether we have good grounds to trust the underlying intuitions that would permit the affluent to neglect the needs of the poor. The critics of maximization may thus do better to shift their focus away from the specific demands of beneficence, and instead emphasize the purely *structural* objections to maximizing consequentialism.

As Railton (1988, 407) notes, "it seems inconsistent with anything like our ordinary understanding of 'morally right' to say that the boundary separating the right from the wrong is to be sharply drawn infinitesimally below the very best action possible. […] 'Wrong' comes into clear application only when we reach actions far enough below normal expectations to warrant real criticism or censure." McElwee (2017, 97) similarly observes that "we do not judge that someone warrants feelings of blame and guilt simply for acting morally suboptimally." It's one thing to hold that extreme poverty and suffering are morally intolerable, and so can generate extreme demands. It's quite another to insist upon the intolerability of suboptimality per se. Imagine getting worked up over a lost penny in Utopia. The view begins to sound literally insane.

Maximization may also face embarrassment when confronted with its structural analogy to the minimalist view on which only the very worst option (among all available) is held to be wrong (cf. Slote 1985, 77). Although less silly-seeming than this opposite extreme,

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

maximization may start to seem less like a well-motivated "default" form for consequentialism to take, and more like a hasty graft of consequentialist ideas upon (p. 501) an incompatible (or at least ill-fitting) base of deontic concepts. For while consequentialists may all agree that a value-maximizing option is *best*, or what we've most moral reason to choose, it's entirely obscure what *further* claim (if any) the maximizer means to make by insisting that the best option is also *obligatory*.[2]

3. Scalar Consequentialism

Suspicions regarding the traditional deontic concepts may naturally lead consequentialists to jettison them in favor of simply evaluating actions on a scale from better to worse. Howard-Snyder (1994, 110) identifies the heart of consequentialism as the claim that "The better a state of affairs, the more moral reason an agent has to produce it." There's no obvious motivation internal to consequentialism for drawing a line between "right" and "wrong" at any particular point on the scale. Whether maximal, minimal, or something in between, the placement of such a line on the scale from best to worst action may seem unacceptably arbitrary. This *arbitrariness* worry motivated the original development of scalar consequentialism in Slote (1985, chap. 5), and also features in the most recent development of the view by Norcross (2020, chap. 2.3, drawing from his 2006). And it is a fair worry as far as it goes. But I think it actually stems from a deeper problem, namely, a lack of clarity regarding what (for consequentialists) any such line is supposed to even *signify*.[3] Once we are clearer on what the line between right and wrong is meant to signify, we may find that the question of where to draw it is more easily answered.

Norcross assumes that deontic binaries would create extra reasons, in a way that's incompatible with consequentialism. Consequentialists should certainly prefer that Joe give an extra $500 to effective charities rather than Jane giving an extra $499, regardless of whether Jane's increment would bump her over the line from "wrong" to "right." But this is just to observe that consequentialists care exclusively about promoting value, and so have no independent concern for the deontic status of an action (Lawlor 2009, 104). We can't conclude from this that actions don't have deontic statuses, since making an evaluative difference is not, in general, a precondition for being a real property.[4] Nonetheless, we may at least wonder, with Norcross, what the significance of an act's (p. 502) deontic status (as right or wrong) is supposed to consist in if it makes no essential difference to what others should do or prefer.

Here I think there are at least two possible answers worth considering. First, if we accept a distinction between moral and nonmoral (e.g., prudential) reasons, we might take an act's deontic status to reflect how these reasons balance out. Along these lines, Lawlor (2009, 106) suggests that beneficent acts might be morally required by a *cost-sensitive* form of consequentialism just when the moral reasons to do the act outweigh the agent's (disproportionately weighted) prudential reasons against. On this view, while some altruistic sacrifice may be required of us, there is some point at which the personal sacrifice becomes sufficiently great that our prudential reasons to favor ourselves trump our moral

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

reasons to promote the impartial good. Portmore (2011, chap. 5) further develops such a *dual-ranking* structure for act consequentialists. On such accounts, the line between right and wrong marks something significant about the agent's overall reasons, without creating extra reasons or otherwise conflicting with core consequentialist tenets.

Alternatively, rather than treating an act's deontic status as a function of the agent's various reasons for action, we might take it to reflect what *reactive attitudes* are warranted by the agent's so acting (McElwee 2010a). I think it is especially natural to understand talk of "obligation" in this way. Thus, for example, Jane's right-making increment could simply have significance for whether she warrants certain negative reactive attitudes, without that giving her any more reason than Joe has to act.

A potential advantage of this approach over dual-ranking is that we aren't committed to distinguishing distinctively moral from nonmoral reasons, or to giving disproportionate weight to an agent's partial interests. A further advantage to understanding obligation in terms of reactive attitudes will emerge in section 4.3. (One could also opt to combine dual-ranking with the reactive-attitudes account of obligation, so while I discuss them separately, I do not mean to suggest that they are strictly exclusive alternatives.)

Of course, as Norcross (2020, chap. 2.4) notes, consequentialists shouldn't understand wrongness in terms of whether others ought to punish or *express* blame toward the agent, as such acts are themselves open to consequentialist assessment, which may diverge radically from our assessment of the original action. Conceivably, it could even be useful in some circumstances to express blame toward someone for acting optimally (perhaps a trickster demon would reward us for behaving so perversely). So the usefulness of expressing blame is just clearly irrelevant to any fair normative assessment of the original action. But this doesn't undermine an analysis of wrongness in terms of blame-*worthiness*, for whether it is useful to express blame is a completely separate matter from whether blame (understood as a negative reactive attitude) is *warranted*,[5] in much the same way that whether it is useful to express fear in some situation is a completely separate matter from whether fear is warranted.

Consequentialists have historically been loathe to acknowledge any such independent norms of fittingness for our reactive attitudes, but it's not clear that this traditional **(p. 503)** aversion is well-motivated. We wouldn't normally think of consequentialism as being in conflict with an epistemologist's claims about our (evidential) reasons for belief, after all. Insofar as there are belief-related actions we can take (e.g., brainwashing ourselves into acquiring a new belief, or verbally expressing some existing belief), consequentialist assessment of those actions is perfectly possible. Nonetheless, whether a belief is rationally warranted, or supported by the evidence, is just a completely separate question from whether it is, in practical terms, worth either inculcating or expressing. But now notice that the same is true of emotions and reactive attitudes. A plausible nonconsequentialist account of when these states are rationally warranted or *fitting* can comfortably exist alongside consequentialist norms for action (Chappell 2012).[6] It may then turn out that we sometimes ought to acquire irrational attitudes, but such "rational irra-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

tionality" is a familiar possibility since Parfit (1984, chap. 1). So I see no fundamental barrier to consequentialists accepting an independent account of blameworthiness.[7] Such an account could in turn ground an account of wrongness that is (contra Norcross) significant without creating extra deontic reasons.

Our discussion so far suggests that the case for (or against) scalar consequentialism really depends upon the larger question of what sense consequentialists can give to a division between right and wrong actions. That's the conclusion that I, at least, am drawn toward. Some, however, have suggested rejecting scalar consequentialism on the independent basis that it fails to be sufficiently action-guiding.

3.1. Practical Guidance

In his original presentation of the idea of scalar morality, Slote (1985, chap. 5.3) raised the worry that by failing to specify which actions are right or wrong, such a view might seem (p. 504) to leave out something essential to a full account of morality. Against this, Slote notes that even nonscalar accounts don't settle all practical questions in circumstances where they give us multiple permissible options. Moreover, Slote suggests, a highly morally motivated agent (who always prefers morally better options over morally worse ones) could find plenty of guidance in the comparative verdicts yielded by scalar morality. It may be true that someone specifically motivated to do *the least that morality requires of him* could no longer be guided by this specific desire. But if this is really the only moral motivation that the agent has, we may be more inclined to judge that the problem here lies with the agent rather than with scalar moral theories.

An alternative response available to the scalar theorist would be to offer a scalar account of right(er) and wrong(er). Sinhababu (2018) suggests that scalar morality can actually offer *richer* guidance than traditional approaches, as it no longer lumps together (e.g.,) mildly bad and truly atrocious options as "equally wrong." While insisting that there is no *fundamental* significance to the dividing point between right and wrong actions, Sinhababu—following Norcross (2005)—offers a contextualist account of rightness as what's better than some salient alternative or amount of goodness. This allows the scalar consequentialist to engage in "rightness"-talk, but in a way that makes the line drawn between "right" and "wrong" merely conventional.

I find it a bit unclear what is achieved by this maneuver. Because this constructed boundary between "right" and "wrong" lacks normative significance, it would seem a mistake for agents to care about it. Insofar as we want fine-grained guidance, this may be found in the scalar theorist's evaluations of options as better or worse (to greater or lesser degrees). Repeating these judgments using deontic vocabulary doesn't seem to add anything. So, rather than offering guidance via scalar "rightness," scalar theorists might do better to simply insist that agents' moral motivation should be responsive to the scalar evaluative differences upon which their theory rests.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Insofar as we feel that something important is left out of the scalar account, a merely conventional reconstruction of deontic language seems unlikely to help. For what those dissatisfied by the scalar approach really want, presumably, is for our moral theory to identify a principled line of *minimal decency* below which we must not fall (on pain of warranting negative reactive attitudes, perhaps). For that, we must move beyond the resources of a purely scalar account of morality.

4. Satisficing

Slote (1984, 140) introduced philosophers to the idea of satisficing consequentialism: "that an act might qualify as morally right through having good enough consequences, even though better consequences could have been produced in the circumstances."[8] That is, rather than insisting that the best outcome be produced, the satisficer identifies a (p. 505) (nonmaximal) value threshold at which outcomes qualify as "good enough." They then affirm that any act is permissible so long as it brings about some such sufficiently good outcome. Only outcomes below this value threshold are impermissible to produce.

Slote's own version of the view is quite radical, licensing what Pettit (1984, 172) calls "unmotivated submaximization": picking a worse option when a better one has been identified and is available at no greater cost. This gives rise to one of the three central objections to satisficing consequentialism: the problem of gratuitous suboptimality.

4.1. Gratuitious Suboptimality

Mulgan (1993, 125) invites us to imagine that, faced with the option to magically save any number of people from poverty, Achilles deliberately chooses a number that is smaller than the total number of impoverished people, insisting that the resulting outcome is "good enough." Achilles' action seems clearly wrong, since he could just as easily have saved a greater number of people from poverty, at no cost. Satisficing consequentialism thus seems at risk of violating a maximally weakened variant of Singer (1972)'s sacrifice principle: If you can prevent something bad from happening, without sacrificing *anything whatsoever*, you ought, morally, to do it!

Bradley (2006) extends the problem by presenting cases in which satisficing consequentialism would seem to allow agents to gratuitously obstruct an optimal outcome that would have occurred without their maleficent interference. But surely no consequentialist should wish to license actions that so gratuitously steer us away from better outcomes.

A natural response for satisficers is to insist that gratuitously suboptimal outcomes fail to qualify as "good enough" in context (Turri 2005). Rogers (2010) develops a complicated form of satisficing consequentialism that meets this desideratum by only permitting suboptimal choices when the morally better alternatives are comparatively costly to the agent. Chappell (2019) similarly argues that the best structure for a satisficing theory to adopt is one of *constrained maximization*, according to which (roughly speaking) agents should do the best they can *without suffering undue burden*. This makes clear that gratu-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

itous suboptimality is never permitted. And it yields a compelling account of supererogation as choosing to surpass the demandingness-moderating constraints in order to achieve even better results. On such an account, the key work for a satisficing theory is to flesh out what counts as acceptable versus undue burdens.

4.2. Arbitrariness

This leads us to what may be the most obvious challenge for satisficers, namely, identifying—on a principled basis—where to draw the line for what counts as "good enough." Is there any nonarbitrary way to do this? As suggested in section 3, I think we can best make progress here by clarifying what we have in mind when asking which acts are good (p. 506) enough—good enough to secure *what status*, exactly? Some form of *minimal decency*, presumably. But as we saw, there are (at least) two ways of fleshing out this idea: (i) directly in terms of our overall reasons for action or (ii) indirectly in terms of warranting reactive attitudes.

If we understand minimal decency in terms of giving sufficient weight to moral reasons (in relation to the weight we give to our nonmoral reasons), we need our complete theory of practical reason to specify what those appropriate weightings are. Call this the "rationalist" view of moral rightness. If there are such varieties of reasons, there's presumably a fact of the matter about their comparative strengths or weights. Satisficers could then appeal to such facts in order to determine, nonarbitrarily, what outcomes are morally "good enough." These outcomes are "good enough" in the sense that any morally better alternatives are subject to countervailing nonmoral reasons that are sufficiently weighty as to rationally justify the agent's refusal to choose them.

Alternatively, on a "sentimentalist" understanding of minimal decency in terms of (say) demonstrating sufficient quality of will as to render the agent blameless,[9] we instead need an account of blameworthiness (or the fittingness conditions for reactive attitudes) to specify the minimum baseline for what counts as adequate moral concern. Once we have such an account, Chappell (2019, 256–57) argues that the satisficing consequentialist can co-opt it to provide a principled specification of how much effort (or burden) morality can require of us. First we ask *how much effort* an agent in certain circumstances must be willing to expend in pursuit of the general good in order to qualify as adequately concerned (according to our independent account of quality of will). Then we can claim that the agent is required to bring about the best results they can *without having to exceed that level of effort* (excessive gains being supererogatory).

Either way, the satisficer's line between right and wrong can be drawn in a principled way. It simply requires drawing upon resources that go beyond the core consequentialist theory: appealing to either nonmoral practical reasons or else fitting attitudes.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

4.3. Options without Constraints

Our final challenge for satisficing consequentialism is more extensional than structural. The worry is that it risks licensing morally atrocious actions. It's a familiar point that, given its rejection of the doing/allowing distinction and associated deontic constraints, traditional act consequentialism is apt to license actions that strike many as intuitively wrong: killing one as a means to saving five, for example. But at least the maximizer has a compelling response available: such acts are necessary for bringing about the best available outcome. However bad it may be for one to die, it would surely be worse for five to do so.

(p. 507) The satisficing consequentialist has no such easy answer. Consequentialism ascribes no essential significance to the distinction between doing and allowing, or between harming and failing to benefit. So if satisficing consequentialism sometimes permits us to suboptimally let others die, it seems that it must equally permit us to suboptimally murder (Mulgan 2001; cf. Kagan 1984, 251). This is a serious problem for the view. After all, however intuitive it may be to say that we are allowed to refrain from saving a life if that would cost us thousands of dollars, that surely isn't worth being stuck with the corresponding verdict that it's permissible to kill someone merely for personal gain.

This problem may be especially pressing for rationalist satisficers, who are committed to their deontic verdicts tracking the agent's overall normative reasons. They may, of course, appeal to typical consequentialist strategies for avoiding counterexamples, for example, noting the risk of worse outcomes if agents felt free to disregard deontic constraints. They might thus recommend that agents adopt a decision procedure that disallows these objectively justified actions. But the mere fact that their view holds suboptimal killings to be objectively justified is arguably disqualifying, regardless of whether they endorse or reject a decision procedure that encourages agents to be guided by this normative fact.

Sentimentalist satisficers may be better placed to weaken the force of the objection in a couple of ways. First, they can note that their deontic verdict of permissibility doesn't entail that the agent has most overall reason to act in this way. It's open to sentimentalist satisficers to insist that agents always have most reason to act optimally, such that suboptimal acts (even when blameless) constitute a kind of *mistake*. So that's something. But it isn't a very satisfying response, as we would ordinarily think that killing someone for (comparatively morally insignificant) personal gain is a paradigm example of a *blameworthy* (and not merely rationally imperfect) act.

A better route for the sentimentalist, I think, is to appeal to features of human psychology that can explain why killing typically reveals a worse quality of will than merely letting die. The relevant psychological facts concern what we find salient. We do not generally find the millions of potential beneficiaries of charitable aid to be highly salient. Indeed, people are dying all the time without impinging upon our awareness at all. A killer, by contrast, is (in any normal case) apt to be vividly aware of his victim's death. So killing tends to involve neglecting much more salient needs than does merely letting die.[10]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Next, note that neglecting more salient needs reveals a greater deficit of good will (Chappell and Yetter-Chappell 2016, 452). This is because any altruistic desires we may have will be more strongly activated when others' needs are more salient. So if our resulting behavior remains nonaltruistic even when others' needs are most salient, that suggests that any altruistic desires we may have are (at best) extremely weak. Nonaltruistic behavior in the face of less salient needs, by contrast, is compatible with our nonetheless (p. 508) possessing altruistic desires of some modest strength—and possibly sufficient strength to qualify as "adequate" moral concern.

Putting these two facts together, then, secures us the result that suboptimal killing is more apt to be blameworthy (and hence impermissible in sentimentalist terms) than comparably suboptimal instances of letting die. It's a neat result for sentimentalist satisficers that they're able to secure this intuitive result without attributing any *fundamental* normative significance to the distinction between killing and letting die.

5. Reconciliation

In the debate between maximizers, satisficers, and scalar consequentialists, we've seen that much hinges on our understanding of what any putative distinction between right and wrong is supposed to signify. Given the various possibilities explored already in this paper, we may wonder whether participants in this debate have always had the same shared concept in mind. That is, we might question whether there is any single determinate "ordinary concept of rightness" for this debate to be about. If ordinary usage vacillates or otherwise underdetermines what is really meant here, the resulting "ordinary concept" may be too amorphous to make sense of the present dispute. As we've seen, it's especially unclear what maximizers have in mind with their "rightness" talk, and whether there's really any substantive disagreement to be found between them and scalar consequentialists.

This diagnosis opens up an attractive new option for resolving the debate in a more ecumenical fashion. We may become *deontic pluralists*, accepting a variety of different deontic concepts (different senses of *right*, *ought*, etc.), and see maximizers, satisficers, and scalar consequentialists as offering complementary insights into different parts of the moral landscape.

5.1. Deontic Monism

To clear the way for deontic pluralism, it will be helpful to assess the rival view that there is a single, privileged sense of rightness. Given the multiple possible ways of constructing deontic terms that we've discussed already in this paper, the only clear basis I can see for insisting upon *deontic monism* would be if one held that there was a *primitive*, indefinable sense of "right" and "wrong," which could take linguistic priority over the various definable senses of these terms. Parfit (2011, 1:165) seems to affirm such a view, using the

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

phrase "*mustn't-be-done*" to express what he calls an "indefinable version of the concept *wrong*."

I think there are good reasons for consequentialists to reject such a view. To begin with, the idea of a primitive property of *mustn't-be-done* seems unacceptably mysterious, in contrast to more familiar normative properties such as *counting in favor* of an action or *rationally warranting* some attitude. Absolutist moral theories such as Kantianism might well make better sense of such a property, so I don't here mean to suggest that it is (p. 509) unacceptably mysterious in some theory-neutral sense that counts against those other theories. I just mean that a primitive notion of *mustn't-be-done* seems mysterious specifically in the context of consequentialism. It seems like a bad fit for the theory. It's not the kind of property that I'd expect consequentialists to comfortably countenance.

Indeed, the arguments for scalar consequentialism make a lot more sense if we take primitive rightness to be their target. The arbitrariness objection returns in full force, since without an analytic connection to *other* normatively significant properties, there would seem no basis for drawing the line between right and wrong at any particular point (maximal or otherwise) on the scale from better to worse acts. Perhaps connections could be restored by accounting for other properties, such as blameworthiness, in terms of primitive rightness. But this would at the very least raise tricky methodological issues about the legitimacy of using the downstream property to fix the location of the upstream one.

Moreover, insofar as I have any grip at all on the concept of *mustn't-be-done*, it seems like it should mark a point of significant discontinuity in the strength of one's *moral* reasons to act, but this is difficult to reconcile with the continuous scale of value that our acts can realize. It also seems like bystanders should be especially concerned to prevent the occurrence of acts that *mustn't be done*. But such claims are incompatible with consequentialism, as we saw in section 3.

Finally, I think the attractiveness of the deontic pluralist package presented next gives us further reason to reject this (less attractive) rival. So let us move on now to presenting the positive view.

5.2. Deontic Pluralism

We may begin by affirming that a scalar account of our moral reasons constitutes the *core* of consequentialism. On this view, the extent to which one has a moral reason to ϕ is purely a function of how good the world would be if one were to ϕ (as compared to other options).[11] The better the outcome, the more reason we have to produce it (Howard-Snyder 1994, 110).

Further, as we saw in section 5.1, scalar consequentialists correctly reject primitive rightness. This is an important insight that (perhaps surprisingly) clears the way for plausible forms of maximizing and satisficing consequentialism to emerge. For while we should reject primitive or indefinable deontic concepts, we may supplement our core scalar view

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

with various *definable* deontic concepts, including ones that are defined in terms that are normatively significant rather than merely conventional.

First we may consider a couple of deontic concepts that are definable in terms of our reasons for action. There's an obvious sense in which we (ideally) morally ought to do whatever we have *most moral reason* to do. Maximizing consequentialism may be most comfortably understood as answering the question of what we ought, in this aspirational sense, to do (Norcross 2020, chap. 2.9). There would seem no reason for any (p. 510) consequentialist to deny the maximizer's view, so understood. More interestingly, some might affirm the hegemonic thesis that this is also what we have *most overall reason* to do (contra dual-rankers and others who posit weighty nonmoral reasons). Addressing what we have most overall reason to do is, after all, much more substantive and interesting than merely addressing some narrower class of "moral" reasons that might yet be outweighed by other considerations. So I would encourage serious consequentialists to defend this more ambitious maximizing view.

Next, as we saw in sections 4.2–4.3, there is a sentimentalist understanding of wrongness in terms of blameworthiness that meshes very nicely with satisficing consequentialism. This allows consequentialists to present a less demanding account of our obligations, and also to account for the intuitive significance of the doing/allowing distinction in an appropriately derivative fashion.

Consequentialists may thus avail themselves of multiple deontic concepts. But you may wonder whether all of them are really needed. Perhaps we can identify one that has normative priority in virtue of its special relevance to first-personal deliberation—that is, as the sense of "ought" that conscientious agents have in mind when they ask themselves, "What ought I to do?"

There are independent grounds for doubting that the deliberative question has a suitably fixed and determinate meaning.[12] Even just focusing on the choice between the ought of most reason and the ought of minimal decency (or blamelessness), we aren't obviously forced in either direction here by the constitutive norms of agential deliberation. Some agents in some contexts are particularly concerned to at least meet the standards for minimal decency, whereas others are more morally ambitious. We can certainly say that it's better for agents to do better. But it isn't clear that there's much more we can say beyond this trivial evaluative observation. In particular, I see no clear basis for insisting that there is just one *proper* aim of deliberation.

On the contrary, I think we can make good sense of why both standards have a limited place in our normative lives. The ought of most reason is perhaps the most obviously significant. It picks out the *best* choice for us to make, the option which is most well-justified, providing an ideal standard to which it makes sense to aspire. (Of course, whether it is practically useful or advisable to aspire to it in any given situation is a further, empirical question. Some may just be disheartened were they to try. But I don't take such prac-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

tical concerns to undermine the in-principle aptness of the aspiration, which is what I'm concerned with here.)

The practical relevance of the ought of minimal decency may be supported in two ways. First, it arguably has more third-personal significance. We properly hold others to account when their actions fall below the baseline of minimal decency and into the realm of the blameworthy. Although it's nice when they do better than the minimum required, we typically don't feel that it's our place to probe too deeply into such matters or to evaluate them too closely. ("How much *exactly* did you give to charity last year?")

(p. 511) Secondly, the baseline of minimal decency may have first-personal significance given our nature as flawed agents who regularly (perhaps even inevitably) do less than the absolute best. Given that we must, practically speaking, make our peace with often failing to meet the ideal standard, it would seem helpful to have a "backup" standard below which we feel we *must* not fall. The ought of minimal decency seems a natural fit for this role. (Again, I don't mean that it is necessarily the empirically most useful guide for us to follow, but just that it makes theoretical sense as a principled basis for ruling out some options as "unacceptably bad" without committing ourselves to acting perfectly.)

All three components of our deontic-pluralist consequentialist package thus strike me as important for an apt moral outlook. It makes sense to aspire to do the best, while recognizing and accepting the reality that, as flawed agents, we will typically fall short. And it makes sense to have a firmer commitment to maintaining a level of at least minimal decency, rather than being willing to plummet to any moral depths without limit. Then, between these two principled standards lies a continuous scale of more-or-less demanding standards that we might choose to target. To help guide us in this choice, we can appreciate that the more good we achieve, the better. But beyond that, there is no authoritative metastandard out there to tell us how high to aim.[13] This observation may be taken to reinforce the scalar "core" of the pluralist view I've defended.

We've seen that deontic pluralism provides consequentialists with the opportunity to accept an attractive package of views: scalar at core, maximizing about the ought of most reason, and satisficing about obligation. Notably, these specific scalar, maximizing, and satisficing claims do not conflict. Moreover, each seems to have a place in our thinking about the normative terrain. So we can, and arguably should, accept all three. Our overall consequentialist theory may be the stronger for it.[14]

References

Anscombe, G. E. M. 1958. "Modern Moral Philosophy." *Philosophy* 33, no. 124: 1–19.

Bradley, Ben. 2006. "Against Satisficing Consequentialism." *Utilitas* 18, no. 2: 97–108.

Chappell, Richard Yetter. 2012. "Fittingness: The Sole Normative Primitive." *Philosophical Quarterly* 62, no. 249: 684–704.

Chappell, Richard Yetter. 2019. "Willpower Satisficing." *Noûs* 53, no. 2: 251–265.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Chappell, Richard Yetter, and Yetter-Chappell, Helen. 2016. “Virtue and Salience.” *Australasian Journal of Philosophy* 94, no. 3: 449–463.

Howard-Snyder, Frances. 1994. “The Heart of Consequentialism.” *Philosophical Studies* 76, no. 1: 107–129.

Kagan, Shelly. 1984. “Does Consequentialism Demand Too Much? Recent Work on the Limits of Obligation.” *Philosophy and Public Affairs* 13, no. 3: 239–254.

(p. 512) Lang, Gerald. 2013. “Should Utilitarianism Be Scalar?” *Utilitas* 25, no. 1: 80–95.

Lawlor, Rob. 2009. “The Rejection of Scalar Consequentialism.” *Utilitas* 21, no. 1: 100–116.

Lazari-Radek, Katarzyna de, and Singer, Peter. 2014. *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. Oxford: Oxford University Press.

McElwee, Brian. 2010a. “Should We de-Moralize Ethical Theory?” *Ratio* 23, no. 3: 308–321.

McElwee, Brian. 2010b. “The Rights and Wrongs of Consequentialism.” *Philosophical Studies* 151, no. 3: 393–412.

McElwee, Brian. 2017. “Demandingness Objections in Ethics.” *Philosophical Quarterly* 67, no. 266: 84–105.

Mulgan, Tim. 1993. “Slote’s Satisficing Consequentialism.” *Ratio* 6, no. 2: 121–134.

Mulgan, Tim. 2001. “How Satisficers Get Away with Murder.” *International Journal of Philosophical Studies* 9, no. 1: 41–46.

Murphy, Liam. 2000. *Moral Demands in Nonideal Theory*. New York: Oxford University Press.

Norcross, Alastair. 2005. “Contextualism for Consequentialists.” *Acta Analytica* 20, no. 2: 80–90.

Norcross, Alastair. 2006. “Reasons without Demands: Rethinking Rightness.” In *Contemporary Debates in Moral Theory*, edited by James Dreier, 38–54. Malden, MA: Wiley-Blackwell.

Norcross, Alastair. 2020. *Morality by Degrees: Reasons without Demands*. Oxford: Oxford University Press.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.

Parfit, Derek. 2011. *On What Matters*. Vol. 1. Oxford: Oxford University Press.

Pettit, Philip. 1984. “Satisficing Consequentialism.” *Proceedings of the Aristotelian Society, Supplementary Volumes* 58:165–176.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Portmore, Douglas W. 2011. *Commonsense Consequentialism: Wherein Morality Meets Rationality*. New York: Oxford University Press.

Railton, Peter. 1988. “How Thinking About Character and Utilitarianism Might Lead to Rethinking the Character of Utilitarianism.” *Midwest Studies in Philosophy* 13, no. 1: 398–416.

Rogers, Jason. 2010. “In Defense of a Version of Satisficing Consequentialism.” *Utilitas* 22, no. 2: 198–221.

Scheffler, Samuel. 1985. “Agent-Centred Restrictions, Rationality, and the Virtues.” *Mind* 94, no. 375: 409–419.

Simon, Herbert A. 1955. “A Behavioral Model of Rational Choice.” *The Quarterly Journal of Economics* 69, no. 1: 99–118.

Singer, Peter. 1972. “Famine, Affluence, and Morality.” *Philosophy and Public Affairs* 1, no. 3: 229–243.

Sinhababu, Neil. 2018. “Scalar Consequentialism the Right Way.” *Philosophical Studies* 175, no. 12: 3131–3144.

Slote, Michael. 1984. “Satisficing Consequentialism.” *Proceedings of the Aristotelian Society, Supplementary Volumes* 58: 139–163.

Slote, Michael. 1985. *Common-Sense Morality and Consequentialism*. London: Routledge & Kegan Paul.

Sobel, David. 2007. “The Impotence of the Demandingness Objection.” *Philosophers’ Imprint* 7, no. 8: 1–17.

Turri, John. 2005. “You Can’t Get Away with Murder That Easily: A Response to Timothy Mulgan.” *International Journal of Philosophical Studies* 13, no. 4: 489–492.

Woollard, Fiona. 2016. “Dimensions of Demandingness.” *Proceedings of the Aristotelian Society* 116, no. 1: 89–106.

Notes:

([1]) Michael Slote has suggested to me a need to distinguish “maximizing” from “optimizing,” as one who gives weight to distributional considerations might judge the optimal distribution to be one that does not maximize the amount of good being distributed. However, once we recognize that the egalitarian consequentialist takes equality itself to be of value, we can see that maximizing the amount of *some resource* without regard for distribution is not necessarily the same as maximizing the *value* of the overall state of affairs. It is the latter that I take the maximizing consequentialist to be committed to. “Maximizing” in this sense is perfectly compatible with egalitarian or other distributional concerns.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([2]) Lazari-Radek and Singer (2014, 334), for example, explicitly take the question "What ought I to do?" to be equivalent to "What do I have most reason to do?"

([3]) Anscombe (1958) famously raised similar worries, albeit in a different dialectical context, about the intelligibility of traditional deontic concepts in modern moral philosophy. Railton (1988, 408) suggests that act utilitarians use "right" as a term of art, which strikes me as robbing their claims about rightness of any substance. Howard-Snyder (1994, 121) aptly observes, "Once you see the consequentialist as saying that there is more moral reason to produce A than B it is hard to see what else she could be saying when she says that the agent *ought to* or is required to produce A."

([4]) We might also question whether a third party's attitudes are the right place to look for moral significance here (Lang 2013, 86). The deontic status of Jane's action may have some significance for Jane even if it is not of interest to others, after all.

([5]) McElwee (2010b, 399) has also emphasized this point in response to Norcross.

([6]) We can (of course) evaluate these states, like anything else, in terms of their contributory value: whether their existence makes the overall state of affairs better or worse, and to what extent. Consequentialists will typically regard this question of value as more important than assessments of rational warrant (we should, for example, prefer that people have useful attitudes than that they have warranted ones). Still, theoretical clarity requires us to recognize the two distinct dimensions of normative assessment here. Consequentialists will hold that the two dimensions coincide in the case of actions: the more useful or worthwhile the act, the more objectively warranted it is to perform. (That's what I mean by "consequentialist norms for action": accounting for our normative *reasons for action* in terms of the value or desirability of so acting.) When assessing mental states, by contrast, we should recognize that the two may come apart. This is because whether a propositional attitude is warranted depends upon whether its propositional content or *object* is *fitting* to the type of attitude that it is, which is clearly different from simply evaluating the consequences of possessing the attitude in question. Note that acts don't allow for any such clear-cut distinction between "state" and "object," which may partly explain why it's so much more natural to think that acts are warranted insofar as they produce desirable outcomes.

([7]) Independent of the consequences of possessing the blaming attitude, that is. There will still be important connections between our normative theory and our account of blameworthiness. Even supposing that blameworthiness is to be accounted for in terms of quality of will, consequentialists are likely to differ from others in how we interpret this. We may judge some beneficent acts of utilitarian sacrifice—e.g., killing one to save five in the notorious *Trolley Bridge* case—to be well-motivated, when a deontologist would disagree.

([8]) Though the broader concept of satisficing first emerged in the economics literature (Simon 1955).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([9]) This is a slight oversimplification. It should be possible to do the right thing for the wrong reasons, after all, and hence act permissibly but in a way that is blameworthy. So the connection must be slightly looser than presented here. It would be better to understand permissible acts as those that are *compatible* with possessing adequate quality of will, given all relevant information. But the precise details aren't essential for our purposes.

([10]) There are exceptions, for example, watching a child drown in a shallow pond right before your eyes—but those are precisely the cases in which we're inclined to judge letting die to be morally comparable to killing.

([11]) Thanks to Doug Portmore for suggesting this formulation.

([12]) See <https://www.philosophyetc.net/2009/06/deliberative-question.html> for discussion of how the relativistic aspects of our assertoric practices here undermine the philosophical significance of the deliberative question.

([13]) Some consequentialists may naturally be drawn to the practical metastandard of asking *what standard is such that our aiming at it would have the best consequences?* That's of some practical interest, but it lacks the special authority of identifying the uniquely fitting or *appropriate* standard to aim at. It is also, notably, not necessarily the *metastandard* such that our employing it would have the best consequences.

([14]) Thanks to Sarah Buss, Anna Edmonds, Emma Hardy, Hrishikesh Joshi, Douglas Portmore, Peter Singer, Michael Slote, and Helen Yetter-Chappell for helpful discussion and comments.

Richard Yetter Chappell

Richard Yetter Chappell is Assistant Professor of Philosophy at the University of Miami. His primary research interests concern the defense and development of consequentialism, effective altruism, and robust normative realism. Chappell blogs at www.philosophyetc.net about these and other philosophical topics. He has published widely in journals, including Noûs, Australasian Journal of Philosophy, Philosophical Studies, and Philosophical Quarterly, and was coawarded the Rocky Mountain Ethics Congress 2013 Young Ethicist Prize.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Global Consequentialism

Hilary Greaves

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.11

Abstract and Keywords

Many types of things are arguably appropriate objects of deontic moral assessment: not only acts but also decision procedures, character traits, motives, public moral codes, and so on. Global consequentialism recommends, for every type that is an appropriate object of deontic assessment at all, that we assess items of that type in terms of their consequences. This (and not simple act consequentialism alone) seems to be roughly the kind of consequentialist thesis that real-life consequentialists, both past and present, have generally been sympathetic to. In this chapter, I articulate a thesis along these lines and defend the thesis in question against the most common objection it faces ("the inconsistency objection"). I discuss the extent to which "going global" deals satisfactorily with three standard objections to act consequentialism: the incorrect verdicts objection, the self-defeatingness objection, and the silence objection. I conclude that global consequentialism has adequate responses to all of these objections, but that it is unclear whether global *consequentialism* is superior to an account that simply stresses the importance of global *axiological assessment*.

Keywords: global consequentialism, axiology, self-defeatingness objection, evaluative focal points, inconsistency objection, actualism, possibilism

1. Introduction

VERY roughly for now, *consequentialism* is the thesis that moral assessment is to be solely in terms of consequences. That statement is a good starting point, but it is very vague. This chapter investigates how best to make it precise, focusing in particular on what kinds of item are to be assessed. Via some discussion of act, motive, and rule consequentialism and other similar theses, I motivate and articulate a form of *global consequentialism*. I then defend the resulting thesis against the most common objection it faces ("the inconsistency objection"). I consider the extent to which global consequentialism ameliorates some of the standard objections to act consequentialism, and briefly compare it to accounts that emphasize axiological over deontic assessment.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The most-discussed consequentialist thesis is

> **Act consequentialism (AC)**: For all choice situations C and all acts x that are available in C, x is permissible in C if and only if there is no alternative act y that is available in C and whose consequences are better than those of x.

Act consequentialism faces at least three well-known objections.

The *incorrect verdicts objection* is that AC simply gives incorrect verdicts in some cases. Suppose, for instance, that a doctor could either treat her patient, or instead harvest that patient's organs for transplant, thereby saving the lives of five others. If we make enough stipulations about the case (for instance, that nobody would ever find out what the doctor did), the consequences of harvesting would be better than the consequences of not harvesting. Even if so, the critics insist, it would be wrong to harvest. Therefore (the objection continues) act consequentialism is false.

The *self-defeatingness objection* stems from the observation that directly consulting the act-consequentialist criterion of permissibility for the purpose of making decisions, (p. 424) on a case-by-case basis, itself would often lead to suboptimal consequences. There are at least four reasons why this observation is correct. First, explicitly performing consequentialist calculations can be a time-consuming business, and time can be precious. (Faced with a child drowning in a shallow pond, the best consequences generally result from *just jumping in to save her*, not from first weighing up the possibility that, say, this child might grow up to be a murderer.) Second, consequentialist calculations involve making nonobvious estimates; as a result, there is significant scope for personal bias to lead one systematically toward incorrect conclusions via attempted explicit calculation, while an appropriately chosen rule of thumb could avoid the systematic error in question. (Faced with a decision between visiting an elderly relative and going to a party, one can easily bring oneself to underestimate how much the visit would cheer the relative. A rule of familial piety would generally lead to better outcomes.) Third, the process of constantly consulting consequentialist considerations is somewhat psychologically unnatural and, as a result, carries a significant psychological cost. (Faced with a decision as to whether or not to visit one's partner in hospital, it is psychologically costly to think in terms of the overall good, rather than simply following the dictates of love.) Fourth, and relatedly, some of the most important goods (such as love) require that the agent be acting on the basis of considerations other than that of the overall good (Hodgson 1967, chap. 2; Smart 1973, 42–44; Stocker 1976, esp. 458–461; Parfit 1984, 27–28; Railton 1984, 134–137). (One's partner might be significantly less comforted by the hospital visit if he knows that it is made as a result of consequentialist calculation, rather than simply from a desire to be with him and to support him.) It is not entirely clear how these facts might lead to an objection to act consequentialism.[1] In particular, nothing in these facts contradicts the content of act consequentialism. Still, perhaps there is some objection lurking here; we return to this later.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The *silence objection* is not an objection to act consequentialism per se, but is rather an objection to the suggestion that act consequentialism covers *all there is to say* by way of moral theory. One might well think that moral theory should also take on such questions as "What motives should I have in this situation?" and "What kind of person should I be?" Since these are not questions about which acts to perform, act consequentialism does not answer them (Stocker 1976, esp. 453, 464–465; Foot 2002, 1).[2]

(p. 425) By way of promissory note: the "global consequentialist" thesis that I will advocate here deals satisfactorily with the second and third of these objections. In addition, it at least dilutes the first objection, despite containing act consequentialism as a proper part.

For related reasons, most modern theorists who are sympathetic to consequentialism at all are (as far as I can tell) sympathetic to something like the type of global consequentialism I discuss in this chapter, and not merely to act consequentialism (for instance: Smart 1973, 48; Hare 1981, chap. 3; Parfit 1984, 25; Railton 1984, 157–160; Pettit and Brennan 1986; Brink 1986, 421; Railton 1988, 410–411; Driver 2001, xiv, 72). Similarly, as an historical matter, global consequentialism is a much more plausible and charitable reading of many of the celebrated consequentialist (specifically, utilitarian) texts, which often articulate ideas along the lines of *global* (rather than act) utilitarianism quite explicitly (Bentham 1823, 102; Austin 1832, 114, 116–120; Mill 1882, 1155–1156; Sidgwick 1907, 405–406, 413). The tendency of modern moral theory to take act consequentialism *alone* as the canonical consequentialist thesis therefore runs a risk of discussing a straw person.

2. Consequentialist Treatments of Decision Procedures and Motives

Ultimately, as we will see in section 3, global consequentialism engages in moral assessment of many types of entity: not only acts but also such things as decision procedures, motives, character traits, laws, public moral codes, and so on. By way of warm-up exercise, we first consider decision procedures and motives.

Suppose that one believes act consequentialism, but also (in partial response to the silence objection, perhaps) wants to engage in moral assessment of decision procedures. There are two obvious consequentialist-spirited ways one might assess decision procedures:[3]

> **$DC_{content}$**: Agents ought to decide what to do by considering which of the available actions would lead to the best consequences.
>
> **DC_{direct}**: Agents ought to decide what to do using whichever decision procedure leads to the best consequences.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

As the "self-defeatingness objection" observes (see earlier), these two criteria often recommend different decision procedures.

We can now articulate one way of turning these considerations of "self-defeat" into an objection. They suggest that given the empirical facts of human psychology, $\mathbf{DC_{content}}$ (p. 426) is at least contrary to the spirit of consequentialism. Someone who takes the spirit of consequentialism to be an exclusive concern with promotion of the good should see little merit in $\mathbf{DC_{content}}$, since $\mathbf{DC_{content}}$ holds that agents ought to decide in such-and-such a way, while deciding in that way tends to lead to worse consequences than some available alternative way of deciding.

Insofar as that is the objection, however, we must note well that it is in fact an objection either to $\mathbf{DC_{content}}$ or to the conjunction $\mathbf{DC_{content}}$ **& AC**, and *not* (as we originally suggested) to **AC** itself. It simply suggests that insofar as a consequentialist wishes to assess decision procedures at all, she should adopt $\mathbf{DC_{direct}}$ rather than $\mathbf{DC_{content}}$. (In fact, **AC** arguably already *implies* $\mathbf{DC_{direct}}$, since arguably "deciding what to do by using decision procedure X" is itself an act.[4])

Let us assume that that point is well taken.[5] The next question concerns the relationship between the favored consequentialist criterion of assessment of decision procedures ($\mathbf{DC_{direct}}$), on the one hand, and the moral assessment of acts, on the other. There are two natural options. First, one could assess acts *indirectly*: hold that $\mathbf{DC_{direct}}$ picks out the permissible decision procedures, and that the permissible acts are the ones that (even if themselves suboptimific) could, or did, result from a permissible decision procedure. Second, one could retain **AC** for the (direct) assessment of acts, alongside $\mathbf{DC_{direct}}$ for the (similarly direct) assessment of decision procedures.

Before discussing this choice between direct and indirect modes of assessment of acts, let us outline a parallel example: the moral assessment of motives (Adams 1976). This is straightforwardly analogous to the moral assessment of decision procedures. Again, one could hold that agents ought to possess the motive that is consequentialist in content ($\mathbf{MC_{content}}$), or one could hold that agents ought to possess whichever motive is such that possession of it leads to the best consequences ($\mathbf{MC_{direct}}$). Again, the spirit of consequentialism favors $\mathbf{MC_{direct}}$. Again, there is an initially open question about which criterion for assessing acts is best combined with $\mathbf{MC_{direct}}$. One could assess acts indirectly (holding that an act is permissible iff it could or did result from a permissible set of motives) or directly (i.e., retaining act consequentialism).

In both cases, the direct option is superior. One reason for this is familiar from discussions of rule consequentialism, which itself includes indirect assessment of acts. As the (standard) "incoherence objection" points out, rule consequentialism threatens to be motivationally incoherent. If the driving idea behind the theory is that morality is about promotion of the good, then it is (not inconsistent, but) motivationally incoherent to hold that some act X is wrong *even when it is known that X would best promote the good*, merely on the grounds that X is forbidden by some rule that *usually* (but not this time) performs better at promoting the good (see Hooker 2016, section 8 and references (p. 427) therein).[6]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Similarly, it would be motivationally incoherent for a consequentialist-in-spirit to hold that an optimific act X is wrong on the grounds that some decision procedure that *normally* (but not this time) better promotes the good would not lead the agent to select X, or that some motive set that *normally* (but not this time) better promotes the good would not motivate the agent to choose X (Smart 1956).

There is a second reason why direct consequentialist assessment of acts is a better partner to such consequentialist theses as $\mathbf{DC_{direct}}$ and $\mathbf{MC_{direct}}$ than is an indirect criterion for act assessment. This second reason stems from the point that indirect assessment needs to tie the evaluation of acts to *one particular* non act type of evaluand to the exclusion of others. One cannot, for example, coherently hold that an act is permissible iff it would be selected by an optimific decision procedure *and* iff an optimific motive set could lead to it *and* iff it conforms to an optimific set of rules, because those three criteria in general contradict one another.[7] But any particular way of resolving this choice seems unacceptably arbitrary. A theory that assesses acts, decision procedures, motives, and rules all in terms of their consequences, on the other hand, does not privilege any type of evaluand over any other, and so does not face this charge of arbitrariness.

3. Global Consequentialism

We are now well on the way to global consequentialism. We have suggested (section 2) that a consequentialist should assess acts, decision procedures, motives, and rules directly, in terms of their respective consequences. The remaining task is to make more precise what the resulting thesis is and to consider how much further the approach should be generalized; that is, which other types of item (if any) should also receive consequentialist assessment.

3.1. Axiological and Deontic Assessment

It is worth pausing to take a step back. Once we have an axiology, we automatically have *one* mode of assessment of any proposition whatsoever: an axiology assigns a value to every possible world, and hence (via the standard machinery of decision theory) to (p. 428) every proposition. Thus, we can assess such diverse entities as acts, motives, and eye colors, all in terms of their value (how *good* or how *morally fortunate* they are; which is the *best or most morally fortunate*, by the lights of the given axiology, from a given range of alternatives). The significance of these assessments is not to be underestimated. It is not a crazy view that axiological evaluations exhaust most of what is important about normative theorizing. I return to this point in the final paragraph of this chapter: it raises a significant question about whether or not global consequentialism is in the end sufficiently well-motivated.

Plausibly, however, for entities of some other types, there is also an important role for moral assessment of a second, distinctively deontic, form. For one very familiar example, plausibly acts are appropriate objects of assessment in terms of *rightness*, and rightness (whether or not it coincides extensionally with bestness) is at least *conceptually* distinct

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

from bestness. It is also plausible (as per the "silence objection" mentioned earlier) that several other types, beyond acts, cry out for assessment that at least conceptually goes beyond the merely axiological. *Decision procedure*, *motive*, and *rule* are quite plausibly examples; others may include laws, systems of government, religions, and diets.

We must not take this too far: it is clearly not the case that *every* type one cares to name is an appropriate object of *deontic* (as opposed to merely axiological) assessment. In the debate over global consequentialism, for instance, some have suggested moral assessment of such entities as climates, eye colors, and rock shapes. But there does not seem to be any role for describing a given climate or eye color as *right* or *permissible*, over and above the observation that it is a *good* one, or the *best* of those available, in a purely axiological sense. And, indeed, the authors who suggest "consequentialist" assessment of climates or eye colors do tend to be the ones who formulate their "consequentialist" thesis in purely axiological terms in the first place (e.g., Parfit 1984, 25; Railton 1988, 409–410).

Let an *evaluative focal point* be a type such that deontic moral assessment is applicable to items of that type (Kagan 1992; 2000). In extensional terms, it is a substantive open question where the boundary lies between those types that are, and those that are not, evaluative focal points in this sense.[8] (Are rules, for instance, appropriately evaluated as anything like right or wrong?) It is also an open question what underlying principles govern the answer to that first question.[9] In addition, it is an open question precisely what is the appropriate concept of normative assessment, in the case of each evaluative focal point. (It is generally accepted that something like *rightness* is the appropriate term of assessment for acts; but it is less clear whether character traits are best described as *right* or instead [e.g.] as *virtuous*.)

It is beyond the scope of this chapter to settle those open questions. In what follows, I will take the notion of an evaluative focal point for granted. I will discuss such items as motives and rules by way of examples, but nothing in the following discussion (p. 429) essentially presupposes that these particular types are indeed among the evaluative focal points. I will also formulate global consequentialism in terms of permissibility, for concreteness, but again very little in the discussion hinges on this choice. (I will flag the sole exception.)

3.2. Roles

Global consequentialism will hold that items of every evaluative focal point should be assessed in terms of "their consequences." For many evaluative focal points, however, the notion of the *consequences of* an item of that type does not immediately make sense. Consider, for instance, rules. One possible rule is a prohibition on lying. The global consequentialist would have us assess that rule in terms of its consequences. But should we take "the consequences of a prohibition on lying" to be the consequences of complete or partial compliance with that rule, the consequences of teaching that rule, the consequences of writing that rule on the walls of the town hall, or what?

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The general point here is that for "the consequences of x" to make sense, we need at least that x is a proposition.[10] *A prohibition on lying* is not a proposition. *That a prohibition on lying is universally complied with*, or *that a prohibition on lying is written on the walls*, are propositions, but they are different propositions. Similar considerations apply to motives, decision procedures, and so on.[11]

The upshot is that to have a well-defined thesis, the global consequentialist needs to relativize evaluations to *roles*. In the earlier example, *rule* is the evaluative focal point, and *being taught, being (completely or partially) complied with*, and *being written on the walls* are roles. Different evaluative focal points correspond to different sets of possible roles: for example, one can comply with a rule but not with an act, and one can perform an act but not a motive. In addition, some roles might in broader terms *make sense* for a given focal point, without that combination of focal point and role forming an appropriate locus of *deontic evaluation.* For instance, it plausibly does make sense to deontically assess which motive one ought to possess, but it does not make sense to deontically assess which motive ought to be spelled out by the first letters of the next twelve people to be born worldwide.

3.3. Formulating Global Consequentialism

This suggests formulating global consequentialism as follows:

(p. 430)

> **Global consequentialism:** For any subject S, any evaluative focal point F, any role R that is applicable to F and any token x of F that is available for role R, it is permissible for S to have x in role R iff there is no alternative token y of F that is available for R such that the consequences of S's having y in role R are better than those of her having x in role R.

Not every global consequentialist will agree with precisely this version of the thesis. First, as noted in passing earlier, some authors will prefer to replace "permissible" with a some other term of deontic assessment (perhaps "right" or "ought" or "fitting").[12] Second, global consequentialism as stated earlier is a form of *maximizing* consequentialism; those who in general (i.e., aside from the project of "globalizing" consequentialism) prefer scalar or satisficing consequentialism to maximizing consequentialism will correspondingly prefer a scalar or satisficing version of global consequentialism. That is, they will want to assert (respectively) a connection between better consequences and deontic superiority, or between permissibility (or whatever) and sufficiently good consequences, rather than a connection between permissibility and the *best available* consequences. These modifications are largely orthogonal to the considerations discussed in this chapter.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

3.4. The Incorrect Verdicts Objection, Again

The original incorrect verdicts objection, as discussed earlier, is an objection to act consequentialism: it seems to many that act consequentialism's verdicts on such cases as that of organ harvesting are just obviously false. A similar objection can be leveled against the global consequentialist's claims regarding evaluative focal points, at least with respect to some hypothetical cases.

To see this, consider, for example, character traits. According to common-sense morality, one ought not to be spiteful. However, a demon could arrange things so that possessing the character trait of spitefulness led (predictably) to better consequences than did possessing any alternative character trait. According to global consequentialism, in a possible world like that, it is permissible (even obligatory) to be spiteful. Those who are attracted by the incorrect verdicts objection to act consequentialism are likely similarly to find this implication of global consequentialism unacceptable. In the case outlined, they will insist, while spitefulness leads to better consequences, still such a vicious character trait is impermissible.

Matters are muddied somewhat by the fact, noted earlier, that (as for many other evaluative focal points) it is not immediately clear what the appropriate term of deontic evaluation is for character traits. Permissibility is not the only option (and nor is it entirely clear what it amounts to, in the case of assessing character). If we take the question of (p. 431) deontic evaluation for character traits instead simply to be that of which character traits are *virtuous*, intuitions against global consequentialism's verdict on the earlier example might be significantly stronger. It seems clearer, perhaps, that spitefulness is not a virtue than it does that it is impermissible to be spiteful even when spitefulness is optimific. (On the other hand, it seems less clear that the question of which character traits are virtuous is properly regarded as a deontic question, as opposed to some other type of evaluation.)

However, even if the question is one of virtue, the global consequentialist's verdict is not obviously implausible. By way of analogy, consider selfishness. One might naively think that selfishness is a vice: that, other things equal, a character is more virtuous the more purely altruistic it is. However, insofar as "invisible hand" arguments show that some degree of selfishness in practice conduces to the overall good (for reasons related to the economy of information), *plausibly* that same degree of selfishness is a virtue. The global consequentialist might further extend this defensive move to launch a corresponding offensive: as is often observed in discussions of virtue ethics, it is unclear what sense or use can be made of the idea that a given character trait is virtuous simply as a matter of primitive, rock-bottom moral fact, without some grounding in terms of how that character trait conduces to some form of good (whether the agent's own flourishing or the overall good). The global consequentialist's verdict on spitefulness is more counterintuitive than the aforementioned (standard) thought about selfishness, but that is easily explained by the remote nature of the demon-controlled possible world in which the result in question holds.[13]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

4. The Inconsistency Objection

4.1. A First Pass

The most-discussed objection to global consequentialism is *the inconsistency objection*. This objection arises because there can be a certain real or apparent tension between global consequentialism's evaluations of items from different evaluative focal points. It (at least apparently) can happen, for instance, that the permissible acts are not ones that would be chosen by any permissible decision procedure, or by an agent who has any permissible set of motives (Adams 1976, 470–475; Parfit 1984, 31–34; Railton 1988, 401–407). Consider the following examples.

(p. 432)

> **The sheriff.** The sheriff has to decide whether or not to execute an innocent scapegoat in order to pacify the angry mob. On this particular occasion, better consequences would result from execution, even when longer-run psychological effects (on the sheriff and on the public) are taken into account. However, the decision procedure that in general has the best consequences includes refusing to even consider executing anyone who one knows to be innocent.[14]
>
> **The loving parent**. Clare could either give her child some benefit, or give much greater benefits to some unfortunate stranger. Because she loves her child, she benefits him rather than the stranger.[15]

In *The sheriff*, it at least seems, global consequentialism would hold that (1) the sheriff ought to execute a scapegoat, (2) the sheriff ought to have internalized a decision procedure that includes refusing to execute scapegoats, and so (3) if the sheriff has internalized the right decision procedure, then she will not perform the right act, and vice versa.[16] Similarly, in *The loving parent*, it at least seems that (1) Clare ought to have motives that include loving and being disposed to favor her child: having those motives does *in general* lead to better consequences, this is not the only decision in Clare's life, and it is not psychologically possible to switch motives instantaneously on demand. But it also seems that (2) in this particular case, benefitting the stranger would lead to better consequences, and again (3) if Clare has the required motives then she will not perform the required act, and vice versa. What should we make of this?

Many commentators take the existence of such tensions to ground a fatal objection to global consequentialism. Hooker, for example, writes that "global consequentialism tells you to use the best possible decision procedure but also not to do the act picked out by this decision procedure. *That seems paradoxical*" (Hooker 2016, section 5, emphasis added; see also Adams 1976, 475–478).

There does indeed seem to be something awkward about the tensions involved in these examples. We must note well, however, that the resulting awkwardness (or "paradox"), so far at least, falls short of contradiction. It would not be *inconsistent*, for instance, to hold

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

that there exist cases in which it is impossible both to perform a permissible act and to have a permissible set of motives.[17] It is therefore worth probing the nature of the alleged problem more deeply.

4.2. Making the Inconsistency Objection Precise

Toward undertaking this task, note that we *do* get an inconsistency from the following (taking *The loving parent* as illustrative example, for concreteness):

(p. 433)

(I1) Clare ought to benefit the stranger.

(I2) Clare ought to love her child.

(I3) Clare's loving her child leads to Clare's not benefitting the stranger.

(I4) For all agents S and all properties A, B: if S ought to possess A and S's possessing A leads S to not possess B, then it is not the case that S ought to possess B.

(I1)–(I4) are jointly contradictory, provided that the unclear phrase "leads to" is interpreted consistently in (I3) and (I4).[18] This unclear phrase lends itself to at least two interpretations, leading to two importantly different versions of the inconsistency objection.

For the first interpretation, let "leads to" be a matter of whether the agent in question *would* perform act Y if she had motive-set X. That is, replace (I3) and (I4) with (I3') and (I4'), respectively:

(I3') If Clare loved her child, she would not benefit the stranger.

(I4') For all agents S and all properties A, B: if S ought to possess A, and [if S possessed A then S would not possess B], then it is not the case that S ought to possess B.

By hypothesis, (I3') is true. (I4'), however, is untenable, for reasons that are independent of any consequentialist commitments: for morally imperfect agents, it can easily be the case (by the lights of any plausible moral theory) that if one satisfied one requirement, one would violate another. (Suppose that Jenny faces some fairly demanding moral requirement in the morning—say, to help her aging and immobile parent get to the bingo and back. Suppose further that if Jenny did this, she would be sufficiently run-down in the afternoon that she will fail to be polite to an awkward stranger at the bus stop. Still, Jenny *ought* both to help her parent and to be polite to the stranger.) Therefore, under this first interpretation, the inconsistency objection is unsound.

For the second interpretation, let "leads to" instead be a matter of whether it would be *possible* for the agent to have motive-set X and yet perform act Y, in some more permis-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

sive sense of “possible.” That is, replace (I3) and (I4) instead with (respectively) (I3’’) and (I4’’):

> (I3’’) It is not possible for Clare to love her child and (yet) benefit the stranger.
>
> (I4’’) For all agents S and all properties A, B: if S ought to possess A and it is not possible for S to [possess A and possess B], then it is not the case that S ought to possess B.[19]

(p. 434) In this case, too, however, considerations independent of global consequentialism force us to reject at least one of the claims (I1), (I2), (I3’’), (I4’’). The details, however, are sensitive to choice points in a background debate about the structural behavior of ought-claims, namely, the debate between actualism and possibilism.[20] We therefore pause to briefly review the key elements of the actualism/possibilism debate, before returning to the inconsistency objection.

4.3. Actualism, Possibilism, and the Inconsistency Objection

The distinction between actualism and possibilism is brought out by examples like the following (Jackson and Pargetter 1986, 235):

> **Professor Procrastinate**. Professor Procrastinate receives an invitation to review a manuscript. Because he is the best-placed person to perform the review, the best possible outcome is that he accepts the invitation and completes the review. However, he is a terrible procrastinator. If he accepted the review, he would not in fact get around to writing the review; it would end up being reassigned to an alternative reviewer, but after some significant delay. That is the worst possible outcome. Alternatively, he could refuse the review assignment, in which case the assignment would go immediately to the alternative reviewer (the second-best outcome).

It is uncontroversial, for present purposes, that in whatever senses it is *possible* for Procrastinate to accept-and-write—that is, to perform the compound action that consists of accepting the assignment and then writing the review—in those same senses he *ought* to accept-and-write. What is controversial is whether Procrastinate ought to *accept the assignment*. *Possibilists* hold that he should accept, on the grounds that he ought to accept-and-write, and accepting the assignment is necessary for performing that compound action. *Actualists* hold that he ought to refuse, on the grounds that the consequences that *would* (rather than: the best consequences that *could*) result from accepting are worse than the consequences that would result from refusing.[21]

(p. 435) Actualism denies (I4’’).[22] To see this, let A be the property of refusing, and let B be the property of accepting-and-writing. Actualism holds that Professor Procrastinate ought to refuse and that he ought to accept-and-write, even though it is not possible for him to possess both properties (refusing, accepting-and-writing). Therefore, this second version of the inconsistency objection, too, is unsound according to actualism. Nor does

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

this seem to be an accident of the particular versions we have considered: the peculiarity that actualism embraces in general seems fundamentally of a piece with the peculiarity that is involved in the inconsistency objection to global consequentialism.

At this point, one might suspect that the inconsistency objection simply presupposes possibilism. That would not obviously refute the objection, since possibilism is (at least) not obviously false. However, in fact possibilists, too, will disagree with at least one of the claims that are required for each version of the would-be objection.

Whether or not (I3′′) is true depends on what sense of "possible" is involved. Presumably, there are some interesting senses of "possible" in which it is possible (although psychologically difficult) for Clare to love her child and yet benefit the stranger, and there are different but also interesting senses of "possible" in which it is not possible for Clare to love her child and yet benefit the stranger. To give the inconsistency objection the best chance of succeeding, let us assume a sense of "possible" that yields the latter verdict, that is, renders (I3′′) true.

If it is not possible for Clare to love her child and yet benefit the stranger, however, then possibilism will not hold both that Clare ought to love her child and that she ought to benefit the stranger. Which properties an agent ought (relative to a given time) to possess, according to possibilism, is a matter of which properties she possesses in the best possible state of affairs that is "available" or "accessible" to her (relative to that time).[23] If, for example, the best available state of affairs is one in which Clare loves her child but does not benefit the stranger, then (I1), (I3′′), and (I4′′) are true, but (I2) is false according to possibilism. Alternatively, if it is better that Clare benefits the stranger at the cost of not loving her child, then (I2) is true but (I1) is false. Either way, again the inconsistency objection is unsound according to possibilism, just as it is according to actualism.

In summary: the phenomenon that gives rise to the inconsistency objection is indeed a prima facie puzzling one, but it arises quite independently from global consequentialism. Further, whatever turn out to be the best resources for dealing with that phenomenon, those same resources will undermine the inconsistency objection.

5. Advantages of Global over Act Consequentialism

Global consequentialism is an extension of act consequentialism. Like act consequentialism, it holds (in maximizing form) that the permissible acts are the ones that lead to (p. 436) optimific consequences. What it adds is that for a potentially large number of other evaluative focal points—plausibly including at least decision procedures, rules, and motives—those focal points, too, are to be deontically assessed directly in terms of consequences.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

In section 1, we briefly surveyed three common objections to (act) consequentialism: the incorrect verdicts objection, the self-defeatingness objection, and the silence objection. Taking these three objections in reverse order, let us now consider the extent to which global consequentialism satisfactorily addresses them.

The silence objection, recall, was that act consequentialism assesses only acts, whereas a complete moral theory should also say something about what motives one should have, what kind of person one should be, and so on. Since global consequentialism has the resources to assess any focal point that is an appropriate locus of deontic assessment, it is clear that no silence objection can apply. The matter of which types of entity are evaluative focal points, and which not, is a choice point that is orthogonal to global consequentialism; the latter can accommodate whichever answer to this question independently seems the most plausible.

The nature of the self-defeatingness objection was never particularly clear. In its naïve form, it is an objection to a certain kind of "consequentialist-flavored" criterion for assessment of various focal points other than acts: an objection to the claim that the right decision procedure involves explicit consequentialist case-by-case calculation, that the right motive is maximization of the good, and so on. (These are the theses that we called $\mathbf{DC_{content}}$ and $\mathbf{MC_{content}}$ earlier, since they build consequentialism into the "content" of the right decision procedure and motive respectively.) Insofar that this is the objection, it is clear that global consequentialism is not vulnerable to it, since global consequentialism eschews those criteria of assessment (instead embracing $\mathbf{DC_{direct}}$, $\mathbf{MC_{direct}}$, etc.).

A more subtle version of the self-defeatingness objection renders it of a piece with the inconsistency objection. Consider again the predicament of Clare. Clare in effect has to "decide" whether to love her child or to benefit the stranger. The consequentialist, global or otherwise, holds that she ought to benefit the stranger. But then global consequentialism perhaps holds in addition that Clare ought to love her child, which, as highlighted by the inconsistency worry, in some sense "defeats" the injunction to benefit the stranger. If this is the worry, however, then we have seen (in section 4) that it is misguided. The tensions involved are not matters of inconsistency, and they arise quite independently of globalization. In fact, they arise quite independently of consequentialism in general: for imperfect agents, it can easily happen that meeting one moral requirement would lead the agent to violate another, on *any* plausible theory of the nature and content of moral requirements.

Finally, one might take the self-defeatingness objection to be the charge that global consequentialism is (true, but) not controversial enough to be interesting. This is perhaps suggested by Bernard Williams:

> [U]tilitarianism's fate is to usher itself from the scene. … [D]irect utilitarianism represents certainly a distinctive way of deciding moral questions, a way, however, which there is good reason to think … could lead to disaster; and some qualifications which (p. 437) [modern utilitarians are] disposed to put in seem to signal some recognition of that, and a comprehensible desire to leave the way open for utilitarianism to retire to a more indirect level. … But once that has started, there

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> seems nothing to stop, and a lot to encourage, a movement by which it retires to the totally transcendental standpoint from which all it demands is that the world should be ordered for the best, and that those dispositions and habits of thought should exist in the world which are for the best, leaving it entirely open whether those are themselves of a distinctively utilitarian kind. If utilitarianism indeed gets to this point, and determines nothing of how thought in the world is conducted, demanding merely that the way in which it is conducted must be for the best, then I hold that utilitarianism has disappeared, and that the residual position is not worth calling utilitarianism. (Williams 1973, 134–135)

As the passage clearly indicates, Williams has in mind an indirect rather than a global form of consequentialism: one which directly assesses only "ways the world is conducted" and indirectly assesses every other focal point in terms of that. However, it is plausible that many of the same considerations apply to global consequentialism. For instance, as we have seen, global consequentialism does hold "that those dispositions and habits of thought should exist in the world which are for the best," and it does "leav[e] it entirely open whether those are themselves of a distinctively [consequentialist] kind." One might well suspect that in this way of thinking, there is little, if any, role for consequentialist evaluation of *acts themselves* to play in moral life, for all that act consequentialism remains officially a part of global consequentialism.

It is a serious overstatement, however, to hold that in the resulting theory "[consequentialism] has disappeared." As is well emphasized by Hare (1981, chap. 3), a conscientious attempt to live one's life in accordance with (anything like) global consequentialism will involve frequent moments of reflection in which one questions whether or not one's existing decision procedures, motives, and so on, are the right ones. *In those moments of reflection*, one explicitly consults the consequentialist calculus and might seek significantly to revise one's decision procedures and motives (etc.) as a result. Further, and relatedly, the recommendations of global consequentialism, in extensional terms, are often distinctive. While global consequentialism is likely to agree with common-sense morality on the status of such rules as "do not lie," "do not kill," and so on, the rules that global consequentialism sanctions are very likely (for instance) to demand significantly more of the global wealthy (such as ourselves) than do the rules of common-sense morality. Nor should this distinctiveness be surprising. The rules (and decision procedures, motives, and so on) of common-sense morality are the result of a number of forces of evolution and power dynamics, and it would be very surprising if they lined up perfectly with the rules (and decision procedures and motives) that maximize the impartial good.[24] The charge that consequentialism has "ushered itself from the scene" is therefore incorrect.

(p. 438) What of the incorrect verdicts objection? A hard-nosed critic of consequentialism would simply insist that, since global consequentialism contains act consequentialism as a proper part, it fares no better than act consequentialism vis-à-vis this objection (and perhaps also adds other similarly incorrect verdicts, concerning the deontic assessment of other evaluative focal points). Global consequentialism continues to hold, for example,

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

that the sheriff ought to execute a scapegoat, and that the doctor ought to harvest organs. For many, this is enough to render the theory untenable.

Before deciding whether this is in the end the right response, however, we would do well to consider the wider range of things that global consequentialism has to say about such cases. For reasons that are well explored in the discussion of self-defeat, global consequentialism will tend to agree (with nonconsequentialism) that the right decision procedure generally refuses even to countenance such acts, that the right rules forbid such acts, that the right character traits are such as to make the agent involuntarily recoil with horror from such acts even if they do recognize that such acts are right, and so forth. The question is therefore whether it is really an inviolable datum that the *acts* in question are wrong, or only that there is *something deeply amiss with a state of affairs* in which the sheriff executes a scapegoat (etc.). Global consequentialism will agree with the latter.

A related point arises from the earlier discussion of Williams. We saw there that it is questionable how much role the assessment of act rightness has to play in moral life, once the global consequentialist's preferred evaluations of various other focal points (such as decision procedures) are in place. In Williams's mouth, this is an objection to the claim that anything is left of consequentialism. But it might alternatively function as a defense of global consequentialism against the charge that the theory's assessment of act rightness is sometimes deeply counterintuitive: if such assessments anyway play at most a highly limited role in moral life, such counterintuitiveness is less objectionable.

Finally, though: it is not clear that anything in the earlier account of the advantages of global consequentialism really draws on the fact that global consequentialism engages in *deontic*, rather than merely axiological, evaluation of such evaluands as motives, character traits, rules, and so on.[25] (Relatedly, it is unclear whether specifically global consequentialism, or instead a view that affords a prominent role to [globally applied] axiology, is the more accurate or charitable reading of the historical utilitarian authors cited in section 1. Since these authors did not in general share the modern moral theorist's concern with distinguishing between axiological and deontic matters, their own writings tend to be somewhat ambiguous on the matter.) The silence objection seems equally well answered by the observation that as soon as one has an axiology, one has verdicts on which character traits, motives, and so forth are better (more morally fortunate) than which others. The incorrect verdicts objection seems equally well answered by the observation that by appealing to axiological evaluation of rules or decision procedures, one can explain why there is something deeply amiss with a state of affairs in which the sheriff executes a scapegoat. Similarly, all the things that global consequentialism says (p. 439) about the self-defeatingness objection have equally good purely axiological analogs. It therefore seems an open question whether the best theory in this vicinity engages in consequentialist deontic assessment of a wide range of focal points, or simply stresses the importance of axiological assessment relative to that of deontic assessment.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Acknowledgments

For useful comments, I am grateful to Richard Yetter-Chappell, Roger Crisp, and Douglas Portmore.

References

Adams, R. M. 1976. “Motive Utilitarianism.” *Journal of Philosophy* 73, no. 14: 467–481.

Austin, J. 1832. *The Province of Jurisprudence Determined*. London: John Murray.

Bales, R. E. 1971. “Act-Utilitarianism: Account of Right-Making Characteristics or Decision-Making Procedure?” *American Philosophical Quarterly* 8, no. 3: 257–265.

Bentham, J. 1823. *An Introduction to the Principles of Morals and Legislation*. Oxford: Clarendon Press.

Brink, D. O. 1986. “Utilitarian Morality and the Personal Point of View.” *Journal of Philosophy* 83, no. 8: 417–438.

Brown, C. 2005. “Blameless Wrongdoing and Agglomeration: A Response to Streumer.” *Utilitas* 17, no. 2: 222–225.

Chappell, R. 2012. “Fittingness: The Sole Normative Primitive.” *The Philosophical Quarterly* 62, no. 249: 684–704.

Driver, J. 2001. *Uneasy Virtue*. Cambridge: Cambridge University Press.

Feldman, F. 1986. *Doing the Best We Can: An Essay in Informal Deontic Logic*. Dordrecht, the Netherlands: Reidel.

Foot, P. 2002. “Virtues and Vices.” In *Virtues and Vices and Other Essays in Moral Philosophy*, 1–19. Oxford: Clarendon Press.

Greaves, H. 2019. “Global Consequentialism and the Morality and Laws of War.” In *Human Rights and 21st Century Challenges*, edited by D. Akande, J. Kuosmanen, H. McDermott, and D. Roser, 59–75. Oxford: Oxford University Press.

Hare, R. M. 1981. *Moral Thinking: Its Levels, Method, and Point*. Oxford: Clarendon Press.

Hodgson, D. H. 1967. *Consequences of Utilitarianism: A Study in Normative Ethics and Legal Theory*. Oxford: Clarendon Press.

Hooker, B. 2016. “Rule Consequentialism.” In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta (Winter 2016 edition). **https://plato.stanford.edu/archives/win2016/entries/consequentialism-rule/**

Hurka, T. 2001. *Virtue, Vice, and Value.* Oxford: Oxford University Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Hursthouse, R., and Pettigrove, G. “Virtue Ethics.” 2016. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 edition). **https://plato.stanford.edu/archives/win2016/entries/ethics-virtue/**

Jackson, F., and Pargetter, R. 1986. “Oughts, Options, and Actualism.” *Philosophical Review* 95, no. 2: 233–255.

Jeffrey, R. C. 1965. *The Logic of Decision*. New York: McGraw-Hill.

(p. 440) Kagan, S. 1992. “The Structure of Normative Ethics.” *Philosophical Perspectives* 6:223–242.

Kagan, S. 2000. “Evaluative Focal Points.” In *Morality, Rules and Consequences: A Critical Reader*, edited by B. Hooker, E. Mason, and D. E. Miller, 134–155. Edinburgh: Edinburgh University Press.

Marcus, R. B. 1980. “Moral Dilemmas and Consistency.” *Journal of Philosophy* 77, no. 3: 121–136.

Mill, J. S. 1882. *A System of Logic: Ratiocinative and Inductive.* 8th ed. New York: Harper & Brothers.

Ord, T. 2009. “Beyond Action: Applying Consequentialism to Decision Making and Motivation.” D.Phil. thesis, University of Oxford.

Parfit, D. 1984. *Reasons and Persons*. Oxford: Clarendon Press.

Pettit, P., and Brennan, G. 1986. “Restrictive Consequentialism.” *Australasian Journal of Philosophy* 64, no. 4: 438–455.

Portmore, D. W. 2011. *Commonsense Consequentialism: Wherein Morality Meets Rationality*. New York: Oxford University Press.

Railton, P. 1984. “Alienation, Consequentialism, and the Demands of Morality.” *Philosophy and Public Affairs* 13, no. 2: 134–171.

Railton, P. 1988. “How Thinking About Character and Utilitarianism Might Lead to Rethinking the Character of Utilitarianism.” *Midwest Studies in Philosophy* 13, no. 1: 398–416.

Sidgwick, H. 1907. *The Methods of Ethics.* 7th ed. London: Macmillan.

Sinnott-Armstrong, W. 1988. *Moral Dilemmas*. Oxford: Blackwell.

Smart, J. J. C. 1956. “Extreme and Restricted Utilitarianism.” *The Philosophical Quarterly* 6, no. 25: 344–354.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Smart, J. J. C. 1973. “An Outline of a System of Utilitarian Ethics.” In *Utilitarianism: For and Against*, by J. J. C. Smart and B. Williams, 3–74. Cambridge: Cambridge University Press.

Stocker, M. 1976. “The Schizophrenia of Modern Ethical Theories.” *The Journal of Philosophy* 73, no. 14: 453–466.

Williams, B. 1965. “Ethical Consistency.” *Proceedings of the Aristotelian Society* 39 (Suppl.): 103–124.

Williams, B. 1973. “A Critique of Utilitarianism.” In *Utilitarianism: For and Against*, edited by J. J. C. Smart & B. Williams, 75–150. Cambridge: Cambridge University Press.

Notes:

([1]) Parfit distinguishes between “direct” and “indirect” forms of self-defeat. According to him,

(i) The general phenomenon of self-defeat is a matter of a theory T “fail[ing] in its own terms, and thus condemn[ing] itself” (1984, 4).

(ii) “Direct self-defeat” occurs when “if someone successfully follows T, he will thereby cause his own T-given aims to be worse achieved” (1984, 55).

(iii) “Indirect self-defeat” occurs if it is true that “if someone tries to achieve his T-given aims, these aims will be, on the whole, worse achieved” (1984, 5).

However, in all these cases, the precise nature of the failure and the resulting condemnation is not immediately clear. Relatedly: the notion of a “T-given aim” is somewhat obscure, and (even if we take that notion for granted) both types of self-defeat fall far short of contradiction.

([2]) There are, of course, acts of *trying to acquire* particular motives, and act consequentialism will issue verdicts on these acts. But one might further think there are moral questions concerning which motives one should *have*. The having of a given motive is not an act.

([3]) Here and elsewhere, I often use “ought” in place of “permissible” for conciseness, but at the cost of some precision (since “ought” has no analog of the distinction between “permissible” and “obligatory”). Relatedly, I will often gloss over the distinction between “best consequences” and “consequences that are at least as good as those of any available alternative,” again for conciseness.

([4]) In this respect, the case of decision procedures is disanalogous to the cases of motives and so on; the latter generate nontrivial work for a global consequentialist thesis to do. I use the case of decision procedures for the warm-up discussion nonetheless, because the

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

treatment of decision procedures within consequentialism has been particularly well worked through in the existing literature.

([5]) It is well taken, for example, by all the authors cited earlier as being "sympathetic to something like global consequentialism." See also Bales (1971).

([6]) In reply to this, Hooker himself (2016, section 8) denies that the motivation for rule consequentialism is a matter of consequentialist foundational principles being intuitively compelling. Instead, he takes the criterion of adequacy to be simply that a moral theory must provide the best available systematization of the particular-case judgments of common-sense morality, and he hypothesizes that rule consequentialism satisfies that criterion. This hypothesis strikes me as implausible, for reasons I hint at later (cf. fn. 23 and related text).

([7]) For example, the optimific motive set might include caring about all sorts of effects that are not publicly traceable to the act, whereas the question of which is the optimific set of rules for public adoption will presumably be more sensitive to the matter of what information about the status of acts is typically publicly available, since here questions of enforceability arise.

([8]) A related fact is that, while ought implies can, the notion of "can" is itself correspondingly unclear.

([9]) Chappell (2012) suggests that the types that are appropriate objects of deontic assessment are all and only the "judgment-sensitive" ones. This suggestion strikes me as plausible, but nothing in the main text hangs on accepting this answer to the question.

([10]) If "consequence" is understood in causal terms, one might further think that some stronger condition applies—perhaps that the evaluand must be an event, in some sense of "event" that does not allow arbitrary propositions (sets of possible worlds) to count as events. For present purposes, however, we can set aside the question of whether any such stronger condition also applies.

([11]) The same considerations apply in principle to acts, but in the case of acts it is perhaps uniquely natural to consider the consequences of the act's being performed (rather than discussed, revered, etc.).

([12]) In addition, as we have seen, some self-styled global consequentialists prefer a purely axiological term such as "best" or "most morally fortunate." However, it is arguably confusing to regard these theses as *consequentialist* theses at all, as opposed to simply (fairly uncontroversial) practices of axiological assessment.

([13]) In the language of Hursthouse and Pettigrove (2016), this is a "virtue theory," to be distinguished from a "virtue ethical theory." A theory of the latter type, by definition, takes the matter of which character traits are virtues to be fundamental, rather than attempting to explain it in terms of anything else. A consequentialist virtue theory along the

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

lines of the one sketched in the main text is defended by Driver (2001, chap. 4). See Hurka (2001, 239–240, 244) for argument against taking the virtues to be fundamental.

([14]) This example could equally be stated in terms of rules, rather than decision procedures.

([15]) This example is from (Parfit 1984, 32). A similar example is Adams' example of Jack and the cathedral at Chartres (1976, 470–471).

([16]) In this section, I use the term "ought to" synonymously with "is required to."

([17]) This is somewhat analogous to the point that while moral dilemmas are puzzling, it is not *inconsistent* to hold that moral dilemmas are possible. Consequentialism notably denies the possibility of moral dilemmas, but others coherently disagree (e.g., Williams 1965; Marcus 1980; Sinnott-Armstrong 1988).

([18]) As I have formulated the inconsistency objection, it insists (in I4) on a connection between arbitrary pairs of properties A, B. This runs some risk of being unsympathetic to the objection: it could in principle be that this completely general principle is false, but that some restricted versions of it (for instance, when A is the property of having some particular motive and B is the property of performing some particular act) are true, and that the restricted versions are enough to ground an inconsistency objection. However, the objections I consider later to the general version (I4) have equally compelling analogs for the various restricted versions that the proponent of the inconsistency objection might propose.

([19]) (I4'') in turn follows from the conjunction of the following two principles:

(Ought-agglomeration): For all properties A, B, (OA&OB) → O(A&B).

(Ought implies can): For all properties A, if OA, then S can possess A.

Ought-agglomeration is questioned by Williams (1965) and Brown (2005).

([20]) My discussion of this point draws heavily on Toby Ord's DPhil thesis (2009).

([21]) More precisely: As is familiar from decision theory, given a probability distribution over possible worlds and a utility function defined on the same set of possible worlds, we can calculate the expected utility of an arbitrary proposition (Jeffrey 1965, 67). Ignoring issues of ties for simplicity, I take *actualism* to be the thesis that S ought to perform A iff the expected utility of A higher than that of any alternative that is available to S. I take *possibilism* to be the thesis that S ought to perform A iff A is part of the maximally specific action that has the highest expected utility, of all that are available to S. Unlike the (standard) formulation in the main text, in terms of what *would* result from each action, this formulation in terms of expected utilities does not presuppose (although it is consistent with) counterfactual determinacy.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

A third position, located somewhere between actualism and possibilism, is *securitism* (Portmore 2011, chap. 6, section 3). Roughly, securitism holds that what S ought to do, relative to time t, is what one does in the best alternative that is *securable* by S at t. Positions in this vicinity are generally regarded as improvements over naïve actualism. For brevity of exposition, I leave this complication aside. Including discussion of securitism would not significantly affect the way in which the actualism/possibilism/(securitism) debate interacts with the inconsistency objection to global consequentialism.

([22]) Relatedly, actualism denies ought-agglomeration.

([23]) See Feldman (1986) for a development of this view.

([24]) For another example of how global consequentialism might lead to distinctive conclusions, see the case study in Greaves (2019). This revisionary character strikes me as a virtue rather than a defect of the global consequentialist approach, although others will disagree (cf., e.g., fn. 5).

([25]) Thanks to Richard Yetter-Chappell for pressing this point.

Hilary Greaves

Hilary Greaves is Professor of Philosophy and Director of the Global Priorities Institute at the University of Oxford. Her main research interests concern issues in moral philosophy, decision theory, and economics, with a special focus on issues that arise in the course of considering how an altruistic actor might most cost-effectively do good. Her published work includes articles on moral uncertainty, population ethics, discounting, and theories of well-being and of interpersonal aggregation.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The Role(s) of Rules in Consequentialist Ethics

Brad Hooker

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.12

Abstract and Keywords

After preliminaries concerning different accounts of the good and the distinction between actual-consequence consequentialism and expected-value consequentialism, this paper explains why consequentialists should prescribe a moral decision procedure dominated by rules. However, act consequentialists deny rules have a role in the criterion of moral rightness. But prescribing a decision procedure dominated by rules and then denying rules a role in the criterion of moral rightness seems problematic. Rule consequentialism gives rules roles first in the decision procedure agents should use and second in the criterion of moral rightness. But giving rules this second role has attracted objections, some of which are outlined and answered here. The final section of the paper considers some recent developments.

Keywords: act consequentialism, rule consequentialism, rules, decision procedure, criterion of rightness, blame, incoherence, reflective equilibrium, collective action problems, partial acceptance problems

CONSEQUENTIALIST ethics is best thought of as a family of theories. The fundamental principles of the theories in this family share a focus on consequences, but some theories in the family evaluate acts solely by their consequences and other theories instead apply the consequentialist test only to other things such as rules or motives (see Portmore, Chapter 1, this volume). The most familiar members of the consequentialist family are act consequentialism and rule consequentialism. The most familiar form of act consequentialism, namely maximizing act consequentialism, holds that an act is morally permissible depending on whether there is some alternative act that would produce better (expected) consequences. Rule consequentialism holds that an act is morally permissible depending on whether the act is permitted by the rules with the best consequences.

The chapter explains why nearly all members of the consequentialist family make use of rules. Rule consequentialism is often accused of giving rules too much importance. That accusation will be assessed here, as will some criticisms of rule consequentialism made by nonconsequentialists. I will also address some recent contributions to the development of rule consequentialism.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Before I turn to rules, I will assemble the building blocks of consequentialism. The next section picks up the concept of the good, as used by consequentialists.

1. The Good

A fairly simple consequentialist theory is utilitarianism, according to which the consequences that matter are additions to or subtractions from aggregate utility. Philosophers typically take "utility" to refer to well-being (i.e., welfare, personal good). There are (p. 442) various views about what is the best account of well-being, such as hedonistic accounts, desire-fulfilment accounts, and objective list accounts (Parfit 1984, 493–502). Here is not the place to explore the contest between these views (Hooker 2015; Fletcher 2016; Crisp 2016; Woodard 2019, chap. 5).

All versions of utilitarianism take the aggregate good to be a matter of the well-being of all, added together *impartially*, such that a benefit or harm to one individual counts for exactly the same as does the same size benefit or harm to anyone else. One of the appeals of utilitarian versions of consequentialism is precisely that they take this impartiality to be essential to the aggregation of utility and thus built into the foundation of morality.

Although we can imagine a consequentialist theory that focuses exclusively on something other than well-being, nearly all consequentialist theories have accorded well-being central importance, if not sole importance. The live question has been not whether well-being matters noninstrumentally but rather whether anything else matters as well noninstrumentally. One answer stretching back perhaps as far as the formula "the greatest good of the greatest number" is that equality matters as well as well-being. Someone who thinks that the consequences that matter noninstrumentally are not only gains or losses in terms of well-being but also equality of well-being could be called a distribution-sensitive consequentialist (Scheffler 1982).

Distribution-sensitive consequentialists might be persuaded that, in addition to aggregate well-being, what matters noninstrumentally is not that everyone has the same level of well-being. There could be situations in which the worse-off cannot be raised to the well-being level of the better-off and yet the better-off could be lowered to the level of the worse-off. Some consequentialists think that, when the better-off are lowered to the level of the worse-off, there is an increase in equality but a decrease in goodness.

These consequentialists might nevertheless be distribution-sensitive. They might hold, for example, that instead of ascribing noninstrumental value to equality of well-being, we should give higher priority to a benefit for someone whose level of well-being is low than to the same size benefit to someone whose well-being is greater. This "prioritarian" view is not committed to "leveling down" the better-off but exerts pressure in favor of raising the worse-off up to higher levels of well-being (Parfit 1997, 202–221; 2012, 399–400). A different form of distribution-sensitive consequentialism holds that what matters noninstrumentally is neither equality nor benefiting the worse off, but instead getting everyone

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

above some threshold of sufficiency (Crisp 2003). Such a view has been dubbed “sufficientarianism.”

Utilitarianism clearly calculates aggregate utility impartially; but do prioritarianism and sufficientarianism? True, such views calculate the good in an agent-neutral manner. In our prioritarian calculation of the overall good, you and I are to give extra weight to benefits to the worse-off compared to the same size benefits to the better-off, whether or not you or I is among the worse-off (see Portmore, Chapter 1, this volume). Likewise, benefits to those below the threshold of sufficiency matter more than benefits to those above, whether or not you or I is below or above that level. But, as I explain elsewhere, impartiality is more than agent neutrality, and so it is not true that a sufficient condition (p. 443) for a consequentialist theory to be foundationally impartial is that the theory be foundationally agent-neutral.[1]

Some recent forms of consequentialism definitely abandon the aspiration to have a theory that is fundamentally agent-neutral. These forms of consequentialism accord more weight to the welfare of *individuals specially connected to the agent* than to the welfare of individuals without such a connection, and more weight to *acts of the agent’s* than to acts of the same kinds performed by other agents. Forms of consequentialism that, at the foundational level, accord more weight to the welfare of individuals specially connected to the agent than to the welfare of others without such a connection and/or accord more weight to acts of the agent’s than to the same kinds of acts performed by others thereby build agent relativity into the foundation of morality.[2]

Consequentialism can have more plausible practical implications if it incorporates into its “value theory” more than just well-being. Consequentialists might even propose that certain kinds of act can have intrinsic *moral* value or disvalue, and that such value must be counted when consequences are assessed.[3] Incorporating into consequentialism the postulation of intrinsic value or disvalue for different kinds of act can help consequentialism have more intuitively plausible consequences. The same is true of building agent relativity into the foundation of consequentialism.

However, each postulation a theory makes, even if the postulation seems intuitively correct, is one more thing the theory assumes rather than explains. If a theory, armed with however many intuitively plausible postulates, can explain the rest of the terrain better than any rival theory, then making those postulates seems necessary in order to come up with the most plausible theory. On the other hand, if some rival theory makes fewer postulates and yet can explain the rest of the terrain as well as the theory that makes a greater number of postulates, then that rival theory has greater explanatory power (since it is explaining just as much but on the basis of fewer postulates). Thus there is what we might call an argument from parsimony against not only forms of consequentialism that postulate intrinsic value or disvalue for different kinds of act but also forms that build agent relativity into the foundation of morality. The argument from parsimony against such forms of consequentialism is that there is some other moral theory that is more parsimonious because it does not start off with such postulates or with agent relativity at the

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

foundational level and yet is just as coherent with our intuitive convictions about right and wrong as such forms of consequentialism are. If there is such a more parsimonious theory, one that has fewer postulates but is equally good at cohering with our moral verdicts, then that more parsimonious theory has considerable advantage.

What reply might come from philosophers who hold that kinds of act do have intrinsic value or disvalue and that there are agent-relative values and disvalues to be (p. 444) incorporated into maximizing act consequentialism? They might say we antecedently believe that, for example, intentionally killing an innocent person has more intrinsic moral disvalue than intentionally letting an innocent person die. Likewise, they might say that we antecedently believe that, for example, your intentionally killing an innocent person has more disvalue for you than do two other people's intentionally killing two other innocent people. If there does appear to be greater intrinsic moral value or disvalue in certain kinds of act, and if there do appear to be agent-relative values to address, then shouldn't we take these appearances as a significant counterweight to the argument from parsimony?

There is disagreement about the correct answer to that question. My own view is that, even if there are such appearances, parsimony seems to me a decisive consideration when we are choosing between two theories that are roughly equally good at cohering with what seems to us to be true. Furthermore, I agree with Woodard (2019, 87) that many claims made about agent-relative value do not seem correct. There does not seem to be more disvalue in your doing something bad than in two other people each doing something equally bad.

This chapter does not have room to argue to a conclusion that there is a more parsimonious theory that can agree with our intuitive convictions about what actions and kinds of actions are right or wrong. The focus in this chapter is on the role(s) of rules in consequentialism. The discussion of the role(s) or rules would become unwieldy if we try to include forms of consequentialism that postulate intrinsic value or disvalue for different kinds of act and forms that build agent relativity into the foundation of morality. Thus, henceforth the chapter will ignore those forms and will concentrate on consequentialist theories that do not postulate intrinsic value or disvalue for different kinds of act and that are completely agent-neutral at the foundational level.

2. Actual versus Expected Value

Another question is whether consequentialism is formulated in terms of *actual* consequences or *probabilities* of consequences (see Cohen and Timmerman, Chapter 7; Bykvist, Chapter 16; and Jackson, Chapter 17, this volume). There is an obvious rationale for caring more about what *actually* happened than about what was *reasonable to predict* would happen. In terms of the impact on well-being, benefits or harms that were possible but not actual matter less than benefits or harms that were actual. And yet at the point of deciding what to do, agents are almost never absolutely certain what all the conse-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

quences of the different actions they might do would be. And so telling them to do what will in fact produce the best consequences is not very helpful.

What is the right way for agents to deal with uncertainty about what the consequences would be of different possible actions? The most familiar answer—at least where the uncertainty is empirical uncertainty about what will happen (as opposed to uncertainty (p. 445) about which moral principles are correct)—concentrates on expected value. Expected value is calculated by multiplying the value or disvalue of each possible outcome times the probability that this outcome would occur and then summing the products. Here is a highly simplified example of an expected value calculation of one item:

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Item being assessed	Values of possible outcomes of item being assessed	Probability of possible outcome	Expected value of possible outcome	Total expected value of item being assessed
An act	One possible outcome of positive value +5	Probability 40%	2	Total expected value of item: 0.2
	Another possible outcome of disvalue –3	Probability 60%	–1.8	

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

An obvious question is: what determines the probabilities? A highly subjectivist answer is that probabilities are determined by what the agent believes about likelihoods. A less subjectivist answer is that probabilities are determined by what people in the agent's milieu believe about likelihoods. And even less subjectivist view is that probabilities are determined by the evidence *available* to people at the time, even if that evidence has not shaped the beliefs held by people at the time.

Whether consequentialism should be framed in terms of actual consequences or expected value remains an unsettled issue. I will return to the distinction between *actual-consequence consequentialism* and *expected-value consequentialism* when I discuss possible conflicts between a prescribed moral decision procedure and an act-consequentialist criterion of moral rightness. However, the next section is about the moral decision procedure that consequentialism prescribes. Choosing the act that actually will produce the best consequences cannot be the prescribed decision procedure, since we typically cannot know which of the acts we might choose actually will have the best consequences. Expected value is more important than actual value "when we are deciding how to act" (Parfit 2017b, 228).

3. The Role of Rules as a Moral Decision Procedure

A self-aware moral decision procedure is composed of a belief about how best to make moral decisions and a corresponding behavioral disposition to choose in the specified way. Since people cannot know the act that actually will produce the best consequences, should people make their moral decisions by calculating the expected values of the alternative acts available and then choosing the alternative with the highest expected value? Nearly invariably, this is not how people should make their moral decisions, for reasons that I will now outline.

(p. 446) Different kinds of ignorance can come into play, and finding out relevant information can be costly or even impossible:

1. Very often, people who have a choice to make may *not know the full range of their available alternatives*. When such information can be obtained, obtaining it typically takes time and attention and imagination.
2. Even when people have this information, they may *not know what the possible consequences are* of some available alternatives. Again, figuring out what the possible consequences are can take time and other resources.
3. Even when people know what the available alternatives are and what each available alternative's possible consequences are, *the value of some of these possible consequences may not be known*. Yet again, figuring out what the value of a possible consequence would be might be costly in terms of time, mental effort, and so on.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

4. Finally, even when people know what the available alternatives are, what their possible consequences are, and what the value of these possible consequences would be, *possible consequences' probabilities may not be known*.

One is not fully equipped to calculate all the expected values of all possible consequences of all available alternatives unless one knows what all the available alternatives are, what all their possible consequences are, what the values of all these possible consequences would be, and what the probabilities are of all the possible consequences. A decision procedure consisting of calculating all the expected values and then choosing the alternative with the highest expected value will be impossible to implement before all the inputs to the calculations are in hand. Sometimes, some inputs cannot be obtained.

Even where all needed information can be obtained, obtaining the information might be very costly. Sometimes the costs of obtaining it are greater than the difference in expected values of various alternatives. Suppose I am deciding between buying inexpensive bicycle A and inexpensive bicycle B for the sake of riding to work more quickly than I can walk. I don't know what the expected value of purchasing A is, and the same is true for purchasing B. But I'm pretty confident the difference in expected values is relatively small. If I spend many months trying to decide which to buy, and I lose forty-five minutes extra time each work day walking to work while I am deciding which to buy, the cost to me of making the decision on the basis of a fully researched expected-value calculation is greater than the difference between purchasing A and purchasing B. Whenever the calculation costs exceed the differences at stake—either by a lot or by a little—a decision procedure consisting of calculating all the expected values and then choosing the alternative with the highest expected value is not cost effective.

The problems I have cited so far with a decision procedure that consists in expected value calculations come from ignorance and the difficulties and costs of overcoming it. These problems would be tremendously important even if, once we had the information needed to do the calculations, we were perfectly accurate calculators. However, we are (p. 447) not perfectly accurate calculators, especially when we have to make practical decisions in a hurry.

In addition, personal biases might distort our assessment either of the values of possible consequences or of the probabilities of those consequences. For example, our bias toward ourselves might lead us to underestimate the *amount* of possible harm to others of decisions that would benefit us. Likewise, our bias toward ourselves might lead us to underestimate the *probability* of such harms.

Moreover, nearly everyone knows personal bias can distort people's calculations in these ways. Now suppose that this knowledge about the influence of bias were combined with a shared belief that the decision procedure used by others for making moral decisions was to try to calculate the expected values of the different alternatives available and then choose the alternative with the highest expected value. If we knew that others' calculations would be distorted by their personal biases and that others would physically hurt us, steal from us, or break their promises to us whenever they convinced themselves that do-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ing such an act has higher expected value than not doing it, we would have little confidence that others would not physically hurt us, steal from us, break their promises to us, or lie. In that case, we would be wise to devote a large share of our attention and other resources to protecting ourselves and our property, and our willingness to embark on diachronic cooperation with others and to trust what others assert would be minimal.

The problems resulting from ignorance, the costs of calculation, the influence of bias, and the need for assurance are each enough to condemn expected value calculation as a decision procedure for everyday use. The aggregate of these problems makes an overwhelming case against such a decision procedure.

Let us refer to the decision procedure the use of which would produce the best consequences, or the highest expected value, as the *optimific* decision procedure. Nearly all consequentialists agree that a decision procedure consisting solely of trying to calculate the expected values of the alternative available acts and then choosing the alternative with the highest expected value would not be the optimific decision procedure. Virtually all consequentialists agree that the optimific decision procedure would consist of dispositions to comply with multiple common-sense rules, including prohibitions and requirements.[4] This decision procedure would be more feasible, cost effective, and reassuring to others than a decision procedure consisting solely of trying to calculate the expected values of the alternative available acts and then choosing the alternative with the highest expected value.

The prohibitions that would most obviously be included in the optimific decision procedure are ones on physically attacking people or their property, stealing, promise-breaking, and lying. But the optimific decision procedure would not consist solely of (p. 448) dispositions to comply with "negative" rules such as the prohibitions just mentioned. Another rule that would be part of the optimific decision procedure would be a "positive" rule about doing good for others in general.

However, this rule of general beneficence cannot be as strong as the rules against assault, theft, promise-breaking, and lying, or we would be back with the increase in danger and distrust that I just mentioned. Furthermore, the rule of general beneficence that is part of the optimific decision procedure cannot require one *always* to maximize expected value as long as one does not infringe the prohibitions on assault, theft, promise-breaking, and lying. If the rule of general beneficence went that far, we would be back with the problems concerning ignorance and getting beyond it that I mentioned earlier.

Should the rule of general beneficence take the form of requiring one to have a standing disposition to benefit others in general when such benefits *are obvious* and providing them would *not involve assault, theft, breaking promises, or dishonesty*? The answer is no, for two different reasons. One of these reasons is that the rule of general beneficence cannot be so categorical that it always trumps duties to those with whom one has special connections (on which, see Jeske, Chapter 12, this volume). The second is that there are limits on the amount of self-sacrifice that the rule of general beneficence can require (see

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Sobel, Chapter 11, and Archer, Chapter 14, this volume). I will explain both of these reasons next.

People often say that parents with an intense interest in general good and little interest in their own children tend to have unhappy children. If what is said here is true, what explains why it is true? Children need love, especially from their parents. Are parents therefore morally required to love their children? Consider this counterargument:

> Premise 1: What can be morally required of one is limited to things over which one has control.
>
> Premise 2: Love is not something over which one has control.
>
> Conclusion: What can be morally required of one does not extend to love.

Both premises of this argument can be challenged. However, I will not do that here. Instead, I will point out that, even if love includes affection and affection is not entirely in one's control, what is in one's control is taking a special responsibility for and interest in one's children. Arguably, even more than affection, what children need is for someone who knows them well to take special responsibility for and interest in them, and to do so on a sustained basis.

If that is what children need, then the optimific decision procedure will include a rule that someone does this for each child. Not always but normally the people most disposed to take special responsibility for and interest in a child are the child's parents. This disposition needs reinforcing by the addition to the decision procedure of an injunction to take special responsibility for and interest in one's own children.

Admittedly, such a rule will often lead people to do what is best for their children when the time, energy, or other resources involved could have instead been used to help other people more. Nevertheless, in general and on the whole, the world will be a less **(p. 449)** miserable place if every child has some people taking special responsibility for and interest in him or her.

I am not at all suggesting that the world would be a happier place if *unrestricted* nepotism pervaded everyday decision making. Nepotism always conflicts with equality of opportunity and typically conflicts with efficiency, and for these reasons it must be restricted. What the restrictions should be on nepotism toward one's children is too big a topic to comment on here. The important point here is that the optimific moral decision procedure is one that requires people to take special responsibility for and interest in their own children.

Although children most obviously need nurturing and protection, nearly all of the rest of us also at least sometimes need affection, attention, and support. Admittedly, some people are loners suited to life away from human contact and attention. But they are exceptions to the generalization that people are social beings. Moreover, arguably, having deep

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

friendships is not only instrumentally valuable to people but also a constitutive element of human well-being.

Central to friendship is mutual affection. A natural concomitant of one's affection for someone is giving that person some degree of priority when one is allowed to do so. Giving priority to friends not only flows from affection for them but also bolsters their affection in turn. A moral decision procedure that forbids prioritizing friends even when one is allocating one's own resources would thus endanger friendship. This is a compelling argument for having a moral decision procedure that *permits* one to prioritize friends when one is allocating one's resources.

We can go further and argue that the importance of friendship is so great that the optimific moral decision procedure would contain a rule *requiring* one to give some degree of priority to one's friends when one is allocating one's resources. One benefit of a requirement to give priority to one's friends is that, as indicated, such priority will help sustain friendships. Another benefit comes from the degree of assurance that internalization of this requirement will give people that their friends will regularly be trying to do good for them.[5]

A comparison between alternative possible rules should not be limited to a comparison of the consequences of people's following the rules. People's acceptance of rules can have consequences that are not the result of actions these people do. For example, many people would find it distressing to be prohibited from taking a special interest in their children, and such distress could predate whatever actions are produced in compliance with the rules. Maybe there are many people for whom the only way to get themselves not to take a special interest in their children would be for them to expunge or at least suppress their affection for their children. For many people, the loss of affection for their children would imperil their relationships with their children and abolish one of the chief sources of purpose and happiness in life. As I suggested earlier, people's loss of affection for their children would also be very bad for their children!

(p. 450) So far, I have pointed to the costs and benefits of being known to accept various rules, the psychological costs and benefits to the agent of accepting various rules, and the difficulties and costs of making decisions by various rules. Another kind of cost I think should be counted are what I call teaching and internalization costs for rules. Teaching and internalization costs are the costs to those who teach the rules and those to whom the rules are taught.

It is possible to learn a rule in the sense of memorizing it while not really understanding the rule's meaning or having one's motivations and dispositions be shaped by the rule. That is not the kind of learning a rule we are interested in here. By far the most important effects of a rule occur if one's motivations and dispositions are shaped by one's understanding the rule. In order to emphasize that the kind of rule acquisition under discussion here involves not merely cognitive belief but also motivational and dispositional ele-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ments, I am using the term “internalization of rules” instead of the term “learning of rules.”

Presumably, the costs (in terms of time and attention and effort) of teaching or internalizing a *greater number* of rules are higher than the costs of teaching or internalizing a smaller number of rules. Likewise, the costs are higher of teaching or internalizing *more complicated* rules rather than less complicated ones. And the costs of teaching or internalizing rules that *conflict with natural inclination* are also higher than the costs of teaching or internalizing rules that require or merely permit actions favored by natural inclination.

To be sure, teaching and learning have their benefits as well as their costs. These benefits must be taken into consideration when we are trying to assess rules or decision procedures. One possible benefit of teaching is the sense of gratification the teacher can get. The benefits of learning are even greater. For example, those learning self-control, trustworthiness, and sensitivity to others’ feelings will probably have their own life prospects improved immensely. Most importantly, an increase in the number, complexity, or demandingness of the rules that get internalized may very well bring benefits to others.

Having identified different costs and benefits of the teaching and internalization of rules, we can run a cost–benefit analysis of different possible sets of rules. Such a cost–benefit analysis will come out favoring a set of rules that includes prohibitions on physically hurting others, stealing, breaking promises, and dishonesty, plus a requirement that one take special responsibility for and interest in one’s children and friends, plus a general requirement that one do good for others when one can do so without physically hurting others, stealing, breaking promises, or dishonesty.

That is not all. There can be cases in which disaster threatens unless one infringes prohibitions on physically hurting others, stealing, breaking promises, and dishonesty, or unless one goes against the interests of one’s children and friends. So, in order to prevent disasters in such cases, the optimific decision procedure would include an overriding prevent-disaster rule (Hooker 2000, 133–136, 165–169).

But there will be limitations on the number, complexity, and demandingness of these rules. The reason for this is that the increased costs of teaching or learning ever more rules, or every more complicated rules, or ever more demanding rules will, at some point, outweigh the increase in the benefits that would be produced by the internaliza **(p. 451)** tion of these rules. Too *many* moral rules are too hard to learn and keep straight. The same is true of too *complicated* moral rules. In contrast, the problem with teaching and internalizing rules that are too *demanding* is not cognitive but motivational. Because people naturally care far more about themselves, their family and their friends than about strangers, getting people to internalize rules demanding self-sacrifice for the sake of strangers is difficult and gets more difficult as the level of likely demands climbs.[6]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

How might such a limit on demandingness be formulated? This limit should be one applied both to the rule about doing good for others and to the rule about preventing disaster. These rules *can* (not must) *require* self-sacrifice over a whole life that is significant. However, except in exceptional circumstances, these rules should not require sacrifice over a whole life that is more than significant. But these rules do *permit* extreme self-sacrifice, even in unexceptional circumstances.

In limiting the amount of self-sacrifice over a whole life that can be morally required so that this level of self-sacrifice need not be more than significant, I admit to being vague. Referring to exceptional circumstances and disasters adds to the vagueness. But a lot of flexibility is needed for limitations on the amount of self-sacrifice required to apply across the full range of possible circumstances. I think that vagueness is the price to be paid for the flexibility needed.

But the vagueness and flexibility are not so great that the limit on demandingness has no bite. Suppose someone asks, Is there a point at which someone has made enough sacrifice for others that she is now justified "to shut the gates of mercy on mankind"?[7] I cannot see how there can be a limit to the demands of self-sacrifice that morality can reasonably require without there being some point where refusing to make further sacrifices is justified.

As I wrote earlier, nearly all consequentialists accept that the optimific decision procedure for making everyday moral decisions is not to try to calculate the expected values of the different possible actions and then to choose the action with the highest expected value. The optimific decision procedure instead contains rules prohibiting certain kinds of action and rules requiring other kinds (with actions that are neither prohibited nor required being optional). While consequentialists typically agree broadly about what prohibitions and requirements are part of the optimific moral decision procedure, I should not leave the impression that all consequentialists agree exactly what the correct test is for decision procedures.

Act consequentialism is *individualist* about this matter in the sense that it prescribes *to you* the decision procedure whose internalization *by you* will produce the best consequences and *to me* the decision procedure whose internalization *by me* will produce the best consequences. Perhaps the decision procedure whose internalization by you will produce the best consequences is not the same as the decision procedure whose internalization by me will produce the best consequences. There might thus be variation between us in how act utilitarianism tells us to make our moral decisions.

(p. 452) Rule consequentialism, in contrast, is collectivist in the sense that it tests a moral decision procedure by the consequences of that decision procedure's internalization *by* everyone. The moral decision procedure whose internalization by the collective would have the best consequences, or has highest expected value, is the one rule consequentialism prescribes to everyone. Rule consequentialism seems to picture morality as a shared, collective enterprise. And rule consequentialism fits smoothly with the idea that morality's requirements, permissions, and prohibitions should be suitable to serve as *pub-*

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

lic rules—to be such that "general awareness of their universal acceptance would have desirable effects".[8]

4. Rules as Part of the Criterion of Moral Rightness?

The term "criterion of moral rightness" is a useful term of art (Bales 1971). A criterion of moral rightness is a complete account of the properties of acts that make acts morally right. Likewise, the criterion of moral wrongness is a complete account of the properties of acts that make acts morally wrong. Rather than refer to criteria of rightness and wrongness, I will abbreviate as appropriate.

Act consequentialism is the view that an act is morally right if, only if, and because of this act's consequences compared to the consequences of alternatives to this act. The most familiar form of act consequentialism requires acts that *maximize* the impartial good, or that *maximize* expected impartial good. A less demanding version of act consequentialism frames the theory in terms of *satisficing*, that is, in terms of bringing about enough good, even if less than the maximum.[9] I henceforth focus on maximizing versions of act consequentialism.

What exactly are the act-consequentialist criteria of rightness? The term "morally right" is ambiguous. The term could mean "morally required" or it could mean "morally allowed" ("morally permitted"). This ambiguity has little importance *within* maximizing act consequentialism. The theory holds that nearly all morally permitted acts are also morally required. The only way, according to act consequentialism, an act, A, can be permitted but not required is if there is at least one alternative act, B, available to the agent in the circumstances that would produce exactly as much good as A and no other act that would produce more good than A or B would. Such circumstances presumably are fairly rare.

(p. 453) When an act consequentialist turns to the question of which decision procedure, rule, or motive is best, there is the question whether the best one is whichever will lead to the greatest number or percentage of good-maximizing acts or whichever will result in the greatest value (Frankena 1988). For the sake of illustration, consider two possible decision procedures, each of which leads to 1,000 acts. The first of these decision procedures would lead to 990 good-maximizing acts but 10 disastrous acts. The second of these decision procedures would lead to 1,000 acts, each one of which is only a little suboptimal. The second decision procedure might well produce greater value even though it produces zero good-maximizing acts.

Of these two versions of act consequentialism, the version that focuses on the greatest value, not the highest number or percentage of good-maximizing acts, seems truest to the spirit of the view. No wonder most act consequentialists take this path.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The distinction between the two versions of act consequentialism I have just been discussing might seem a point about theory rather than practice. These two versions of act consequentialism probably agree *in practice* on what the elements of the optimal decision procedure are. They will agree on this if they think the very same elements will both produce the greatest value and result in the higher number and percentage of good-maximizing acts.

One thing that all versions of act consequentialism agree about is that the role that rules appropriately play is in the recommended decision procedure, not in the criterion of moral rightness or in the criterion of moral wrongness. The act-consequentialist criterion of moral rightness makes no reference to rules: an act is morally right if, only if, and because no other act has higher value. Act consequentialism holds that one can use the morally best decision procedure and yet be led by it to select a morally wrong act, as I will now illustrate.

Suppose you are asked by your boss whether you left work early yesterday. The decision procedure that consequentialism prescribes includes a rule against lying. Admittedly, it also includes a rule requiring you to prevent disaster, and that rule will sometimes conflict with and outweigh the rule against lying, but only where the only way to prevent disaster is to lie. Suppose that in the case at hand you cannot see that lying to your boss is needed to prevent disaster. Thus, you tell the truth that you left early from work yesterday.

Now suppose that your telling the truth to your boss in this case does not maximize the impartial good, because the truth that you reveal annoys your boss and creates tension between the two of you. So the decision procedure that consequentialism tells people to follow leads you to do something that does not actually maximize the impartial good. I mentioned earlier that consequentialist theories can be formulated in terms of actual consequences or in terms of expected value. According to actual-consequence act consequentialism, your act of truthfully answering your boss's question is not morally right, because it does not, as things turn out, produce as good consequences as your lying would have.

Is your telling your boss the truth about your leaving work early yesterday morally right according to expected-value act consequentialism? We earlier discussed many (p. 454) impediments to the calculation of expected value. Definitely, in the few seconds between your boss's asking you whether you left work early yesterday and your having to answer or refuse to answer, you couldn't conduct an expected value calculation that takes into account every possible consequence of telling the truth, every possible consequence of lying, and every possible consequence of refusing to answer. But let us set all these difficulties aside and imagine that you could in fact quickly calculate the expected values of telling the truth, lying, and refusing to answer. And suppose you know your boss and your coworkers were elsewhere yesterday afternoon and so couldn't have seen that you left early. Thus, you are very confident that you wouldn't be found out if you told a lie now. Suppose you are also nearly certain that your boss's hearing the truth will annoy her and

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

create tension between the two of you. Thus, your expected value calculation comes out in favor of lying to her.

As I explained, consequentialism advocates a moral decision procedure dominated by multiple rules. Where the expected values of acts that comply with these rules are lower than the expected value of acts that infringe these rules, the prescribed moral decision procedure leads to an act that is morally wrong according to expected-value act consequentialism. In such cases, the fact that rules play a role in the decision procedure that act consequentialism prescribes but not in expected-value act consequentialism's criterion of rightness yields practical conflict between the prescribed decision procedure and the theory's criterion of rightness.

Admittedly, conflict between the prescribed decision procedure and the criterion of rightness is likely to arise less often for expected-value act consequentialism than for actual-consequence act consequentialism. There are many cases where an act that violates the rules is committed but is not found out. In many of these cases, the consequences of the violation *actually* are better than the consequences of not violating the rule would be. However, when an act that violates the rules is found out, the consequences are typically extremely negative, including blame, loss of trust, withdrawal of good will toward the perpetrator, and perhaps punishment. Even if the probability of the perpetrator's being caught is low, the negative consequences of being caught are typically so bad that an expected value calculation of violating the rule will often come out against violating the rule. There are plenty of instances of rule violations that *turned out* to maximize the impartial good but that an expected value calculation done *in advance* would have opposed.

Rule consequentialism is the view that an act is morally wrong if, only if, and because it is forbidden by the code of rules whose widespread internalization would produce the greatest expected value. Rule consequentialism broadly agrees with act consequentialism about the kind of moral decision procedure people should have, namely that it is one in which multiple rules predominate. However, unlike act consequentialism, rule consequentialism holds that multiple rules have an ineliminable role to play in the criteria of moral rightness and wrongness. According to rule-consequentialist criteria, acts are right or wrong depending on whether they are forbidden by rules that pass a consequentialist test.

Concerning many situations, rule consequentialism and act consequentialism will be in agreement about which available acts would be morally required, morally optional, or morally wrong. Whenever act consequentialism holds that telling the truth or keeping a (p. 455) promise or leaving other people's property alone would maximize the good, rule consequentialism agrees that such acts are morally required. Nevertheless, act consequentialism and rule consequentialism disagree about *why* such acts are morally required. Act consequentialism maintains that they are morally required simply because these *acts* maximize the good. Rule consequentialism holds that these acts are morally required because the optimific *rules* require them.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

5. A Problem with Having the Decision Procedure and the Criterion of Rightness Conflict

Problems for act consequentialism arise from the possible conflicts between its criterion of rightness and its decision procedure. On the occasions where agents follow the decision procedure that act consequentialism prescribes but the act selected by this decision procedure is wrong according to the act-consequentialist criterion of moral wrongness, what judgement does act consequentialism reach? Perhaps that seems like a misguided question. After all, the question itself states both that the agents follow the decision procedure act consequentialism prescribes and that the act is wrong according to act consequentialism. But such a statement seems to pose the question of whether act consequentialism can—without an air of paradox—condemn an act that was selected by precisely the decision procedure that act consequentialism tells agents to use. Surely, to maintain that an act is morally wrong is to condemn it. Yet should an act be condemned if the agent who chose it followed the appropriate procedure in choosing it?

The act-consequentialist response to such queries is to insist that different questions get different answers. Which decision procedure should be used? The act-consequentialist answer is the decision procedure which produces the most good. What acts are morally right? The act-consequentialist answer is whichever ones actually maximize the good or at least have the highest expected value.

Should blame be directed at agents who faithfully follow the prescribed decision procedure and yet choose an act that turns out not to maximize the good or an act that does not have the highest expected value? Act consequentialists point out that this question about blame is different from the question of whether the agent followed the prescribed decision procedure and different from the question of whether the act thus selected maximized value. And the answer that act consequentialists usually give is that an act or agent should be blamed if and only if blaming the act or the agent will maximize the good.[10]

(p. 456) The combination of the act-consequentialist thesis about how agents are supposed to make their moral decisions, the act-consequentialist thesis about which acts are right or wrong, and the act-consequentialist thesis about when blaming agents is appropriate is an extremely counterintuitive combination. Suppose an agent follows the decision procedure that act consequentialism tells the agent to use and, on this basis, decides to do X. Suppose X happens also to be the morally right act in the circumstances, according to the act-consequentialist criterion of rightness. Still, act consequentialism maintains that whether it is appropriate to blame this agent for doing X is an open question. This combination of theses is not inconsistent within act consequentialism. Nevertheless, the combination is extremely counterintuitive.

Should act consequentialists move from the view that whether it is appropriate to blame an agent depends upon whether blaming the agent has the best consequences to the view that whether it is appropriate to blame an agent depends upon whether the agent, without excuse, did something wrong according to act consequentialism? This move would do

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

nothing to address the possibility that an agent could faithfully employ the moral decision procedure that act consequentialism prescribes and yet do what is wrong according to the act-consequentialist criterion of wrongness.

Another way of responding to the problem of possible conflicts between the moral decision procedure and the criterion of wrongness is to amend the criterion of wrongness so that it more closely matches the prescribed moral decision procedure. Since rules predominate the prescribed moral decision procedure, likewise they would have to predominate the criterion of wrongness. But to respond in this way is to abandon act-consequentialist criteria of rightness and wrongness and move to rule consequentialism. Rule consequentialism takes rules to predominate the criterion of rightness, as we saw at the end of section 4.

6. Is Taking Rules To Be Part of the Criterion of Rightness a Mistake?

The foundation of rule consequentialism is its principle about deontic status: whether an act is morally required, optional, or wrong depends on what acts are required, permitted, or forbidden by the rules with the best consequences. So the foundational principle of rule consequentialism builds rules into the criteria of rightness and wrongness.

Consider a case where doing what is required by a rule selected for its consequences happens to be an act that neither has as great expected value as some other act nor would actually have the best consequences. An example might be an act of keeping a promise. If rule consequentialism nevertheless tells the agent to keep the promise, is rule consequentialism being true to its nature as a consequentialist theory? If what all forms of consequentialism care most about is making the consequences as good as possible, then any form of consequentialism that requires an agent to comply with some rule when this (p. 457) act would not produce the best consequences is incoherent.[11] This implies that the act-consequentialist criterion of rightness is the only criterion of rightness that coheres with consequentialism.

This objection might be telling if our argument for rule consequentialism started with the premise that what is most important is maximizing impartial good. If rule consequentialism sometimes reaches a conclusion about which act is morally required where this act is not the one that would maximize impartial good, then that conclusion would indeed fail to cohere with a starting premise that what is most important is maximizing impartial good. But we might have an argument for rule consequentialism that does not begin with the consequentialist premise that what is most important is maximizing impartial good. Indeed, we might have an argument for rule consequentialism that contains no consequentialist premise at all. That would be desirable, since any argument for rule consequentialism that employs a consequentialist premise would strike nonconsequentialists as question begging.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

I should stress that I am distinguishing between the foundational principle of rule consequentialism and an argument for rule consequentialism. The foundational principle of rule consequentialism does indeed assess acts by rules that are selected for their consequences. This principle is consequentialist in an obvious sense. The argument for rule consequentialism, however, is dialectically prior to that foundational principle. And this argument might have no consequentialist premise.

One prominent argument for rule consequentialism is that it does better than any other moral theory at supplying a foundational impartial principle that achieves a reflective equilibrium with our more specific moral intuitions about what kinds of act are right or wrong and about what particular acts are right or wrong.[12] This reflective equilibrium argument for rule consequentialism has no consequentialist premise.

Having seen that there is an *argument* for rule consequentialism that has no consequentialist premise, much less a premise that makes an overriding commitment to maximize the good, we now should consider whether a rule-consequentialist *agent* must have an overriding commitment to maximize the good. I indicated earlier that one rule a rule-consequentialist agent should accept is a requirement to do good for others in general if possible, at least up to some limit of self-sacrifice over a whole life. But this rule does not have overriding force. If it did have overriding force, we would be back with most of the problems we saw with having maximizing the good as one's decision procedure.

Another rule that a rule-consequentialist agent should accept is one to *prevent disasters* if possible, at least up to some limit of self-sacrifice over a whole life. The inclusion of a rule (p. 458) requiring the agent to prevent disasters is motivated on rule-consequentialist grounds (think of the benefits produced when people comply with this rule). This rule springs into operation when disasters threaten, even when preventing the disaster is possible only if the agent breaks some other rule. For example, if breaking a promise is necessary in order to prevent a disaster, then rule consequentialism tells the agent to break the promise.

A different objection from the one that rule consequentialism is incoherent is the objection that rule consequentialism reaches implausible verdicts about what to do. Suppose that the only way to prevent some disaster—such as a death or long-lasting misery—would be for the agent to break a promise of less than momentous proportions. In this sort of case, any moral theory that insisted that the promise be kept is counterintuitive. If rule consequentialism compels the agent to comply in such a case with the rule against breaking promises, then the theory is counterintuitive. Such an objection to rule consequentialism can come from act consequentialists, whose own theory would not require an agent to keep a promise when breaking it would produce better consequences. The objection can also come from those nonconsequentialists who are not absolutists about promise keeping.

However, this objection to rule consequentialism is misguided. The inclusion of an overriding prevent-disaster rule in the set of rules prescribed by rule consequentialism inocu-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

lates the theory against the objection that it would insist on compliance with rules even when disaster results.

The overriding commitment to *prevent disasters* must not be confused with an overriding commitment to *maximize the good*. There are cases where breaking a promise, telling a lie, stealing, or another kind of act condemned by one of the rules would produce slightly better consequences than not doing so would. In such cases, an overriding requirement to maximize the good would insist that the other rule be broken. An overriding commitment to prevent disasters would not have the same result, apart from in the extremely exceptional cases where breaking the other rule would prevent disaster and not breaking it would not, even though the difference in how good the consequences would be of the two acts would be only slight.

The objections that rule consequentialism must be incoherent and that this theory might get in the way of preventing a disaster are therefore misguided. These two old objections to rule consequentialism do not provide sufficient grounds for rejecting the rule-consequentialist position that assigns to rules a central role not only in the moral decision procedure agents should use but also in the criteria of rightness and wrongness. That is not to say that the rule-consequentialist position is not mistaken for some other reason, either old or new.

7. Recent Developments of Rule Consequentialism

Another old objection to rule consequentialism is that it makes implausible demands on the agent when others are not following the optimific rules. Suppose one of the optimific (p. 459) rules requires all of us to restrict our behavior in a certain way for the sake of the environment. Suppose doing so is burdensome to the individual, although of course everyone is better off if the environment is preserved than if it is spoiled. Now suppose you notice that others are in fact not restricting their behavior. The objection is that rule consequentialism requires you to follow the optimific rule even when this will be costly to you and do no good.

The objection that such a requirement is unfair was made by David Lyons (1965, 128–132, 137–142). The objection has been given a new twist and directed at Derek Parfit by Douglas Portmore (2017). Portmore contends that where others are not following optimific rules that are burdensome to the individual agent to follow, and where the individual agent’s following these rules would not compel others to start following them and would not produce some other good, then there would be no sufficient reason for the agent to comply with these rules. If that is correct, then, in such a case, rule consequentialism would be requiring the agent to do something that she has no sufficient reason to do.

This line of objection is very important, and perhaps especially as directed at Parfit’s final position. Parfit’s 2011 volumes argued that there are plausible forms of contractualism

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

that converge with rule consequentialism in claiming that optimific rules determine right from wrong. Parfit's 2017a lays far more emphasis on two other arguments for rule consequentialism. One is the reflective equilibrium argument (2017a, 421–422, 433, 450). The other is an argument about what explains the wrongness of acts that harm very many but harm each to only a very small or even imperceptible degree. This argument appeals "not to the separate effects of particular acts, but to the combined effects of what we and others together do. Some act would be wrong, we believe, if all optimific rules would condemn such acts" (2017a, 432).[13] If Parfit ended up having as one of his two main arguments for rule consequentialism that the theory handles cases where collective action is needed to prevent terrible aggregate harms, then his version of rule consequentialism had better not make implausible demands in such cases. Parfit (2017b, 227, 229) replied to Portmore by contending that rule consequentialism should be formulated such that the rules have to be optimific across different rates of acceptance by the population, and that one such rule would be "do not make sacrifices when these acts … would do no good whatever."[14]

Another recent development in rule consequentialism can be found in essays by Susan Wolf and David Copp. Wolf outlines conceptualizing morality as a practice, "that is, as a loose and informal institution, itself perhaps embracing some smaller sub-institutions, constituted by a set of rules that specify 'offices, roles, moves, penalties, defences,' and so on" (Wolf 2016, 138). If one does think of morality in this way, then "one may without inconsistency or rule-worship admit that the point of the practice of morality is to bring about the greatest good for the greatest number without (p. 460) being committed to the idea that this is also the point of one's life." And Copp (2020) conceptualizes the object of morality as being "to lead the society's members to be disposed such that they can live together with a minimum of conflict, meeting their needs and pursuing their values, cooperating with each other in joint projects that are important to them." If that is the point of morality, then Copp thinks an ideal moral code is one the currency of which would achieve this object at least as well as the currency of any alternative code.

The conceptualization of morality as a practice with the object Copp outlines is one that I accept. And I agree with Wolf and Copp that this conceptualization of morality goes hand in glove with rule consequentialism. However, we cannot argue for rule consequentialism on the basis of this fit without first providing a compelling argument for conceptualizing morality in the way that Wolf and Copp do. The way to mount that argument seems to me to return us to the reflective equilibrium methodology. This conceptualization of morality fits better with our general and more specific beliefs about morality than rival conceptualizations do, or so I intend to argue in a forthcoming paper.[15]

References

Baier, K. 1958. *The Moral Point of View*. Ithaca, NY: Cornell University Press.

Bales, R. E. 1971. "Act-Utilitarianism: Account of Right-Making Characteristics or Decision-Making Procedure?" *American Philosophical Quarterly* 8: 257–265.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Brandt, R. 1989. “Morality and Its Critics.” *American Philosophical Quarterly* 26: 89–100.

Copp, D. 2020. “The Rule Worship and Idealization Objections Revisited and Resisted.” In *Oxford Studies in Normative Ethics* 10.

Crisp, R. 2003. “Equality, Priority, and Compassion.” *Ethics* 113: 745–763.

Crisp, R. 2016. “Well-Being.” In *Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. **https://plato.stanford.edu/entries/well-being/**

Feldman, F. 1992. *Confrontations with the Reaper*. New York: Oxford University Press.

Feldman, F. 1997. *Utilitarianism, Hedonism, and Desert*. New York: Cambridge University Press.

Fletcher, G. 2016. *The Philosophy of Well-Being: An Introduction*. London: Routledge.

Frankena, W. 1988. “Hare on Levels of Moral Thinking.” In *Hare and Critics*, edited by D. Seanor and N. Fotion, 43–56. Oxford: Oxford University Press.

Gert, B. 1998. *Morality*. New York: Oxford University Press.

Hare, R. M. 1981. *Moral Thinking*. Oxford: Oxford University Press.

Harrod, R. F. 1936. “Utilitarianism Revised.” *Mind* 45: 137–156.

Hooker, B. 1995. “Rule-Consequentialism, Incoherence, Fairness.” *Proceedings of the Aristotelian Society* 95: 19–35.

(p. 461) Hooker, B. 2000. *Ideal Code, Real World*. Oxford: Oxford University Press.

Hooker, B. 2007. “Rule-Consequentialism and Internal Consistency: A Reply to Card.” *Utilitas* 19: 514–519.

Hooker, B. 2008. “Rule-Consequentialism versus Act-Consequentialism.” *Politeia* 24: 75–85.

Hooker, B. 2010a. “When Is Impartiality Morally Appropriate?” In *Partiality and Impartiality: Morality, Special Relationships, and the Wider World*, edited by B. Feltham and J. Cottingham, 26–41. Oxford: Oxford University Press.

Hooker, B. 2010b. “Publicity in Morality: Reply to Katarzyna de Lazari-Radek and Peter Singer.” *Ratio* 23: 111–117.

Hooker, B. 2013. “Egoism, Partiality, Impartiality.” In *Oxford Handbook on the History of Ethics*, edited by R. Crisp, 710–728. Oxford: Oxford University Press.

Hooker, B. 2015. “The Elements of Well-being.” *Journal of Practical Ethics* 3: 15–35.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Hooker, B. 2016. “Wrongness, Evolutionary Debunking, Public Rules.” *Etica & Politica* 18: 133–149.

de Lazari-Radek, K. and Singer, P. 2014. *The Point of View of the Universe*. Oxford: Oxford University Press.

Lyons, D. 1965. *Forms and Limits of Utilitarianism*. Oxford: Oxford University Press.

Mason, E. 1998. “Can an Indirect Consequentialist Be a Real Friend?” *Ethics* 108: 386–393.

Mill, J. S. 1861. *Utilitarianism. Fraser’s Magazine* 64: 391–406, 525–534, 659–673.

Mulgan, T. 2001. *The Demands of Consequentialism*. Oxford: Oxford University Press.

Mulgan, T. 2006. *Future People*. Oxford: Oxford University Press.

Parfit, D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

Parfit, D. 1997. “Equality and Priority.” *Ratio* 10: 202–221.

Parfit, D. 2011. *On What Matters*. Vols. 1 and 2. Oxford: Oxford University Press.

Parfit, D. 2012. “Another Defence of the Priority View.” *Utilitas* 24: 399–440.

Parfit, D. 2017a. *On What Matters*. Vol. 3. Oxford: Oxford University Press.

Parfit, D. 2017b. “Responses.” In *Reading Parfit*, edited by S. Kirchin, 189–236. Abingdon, UK: Routledge.

Pettit, P. 1997. “The Consequentialist Perspective.” In *Three Methods of Ethics*, edited by M. Baron, P. Pettit, and M. Slote, 92–174. Malden, MA: Blackwell.

Portmore, D. 2011. *Commonsense Consequentialism: Wherein Morality Meets Rationality*. New York: Oxford University Press.

Portmore, D. 2017. “Parfit on Reasons and Rule Consequentialism.” In *Reading Parfit*, edited by S. Kirchin, 135–152. Abingdon, UK: Routledge.

Powers, M. 2000. “Rule Consequentialism and the Value of Friendship.” In *Morality, Rules, and Consequences*, edited by B. Hooker, E. Mason, and D. E. Miller, 239–254. Edinburgh: Edinburgh University Press.

Rawls, J. 1951. “Outline for a Decision Procedure in Ethics.” *Philosophical* Review 60: 177–197.

Rawls, J. (1971). 1999. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Scheffler, S. 1982. *The Rejection of Consequentialism*. Oxford: Oxford University Press.

Sidgwick, H. 1907. *Methods of Ethics*. 7th ed. London: Macmillan.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Slote, M. 1984. “Satisficing Consequentialism.” *Proceedings of the Aristotelian Society Supplementary Volume* 58:139–163.

Slote, M. 1985. *Common-Sense Morality and Consequentialism*. London: Routledge and Kegan Paul.

(p. 462) Slote, M. 1989. *Beyond Optimizing*. Cambridge, MA: Harvard University Press.

Tobia, K. 2013. “Rule Consequentialism and the Problem of Partial Compliance.” *Ethical Theory and Moral Practice* 16: 643–652.

Urmson, J. O. 1953. “On the Interpretation of the Moral Philosophy of J. S. Mill.” *Philosophical Quarterly* 3: 33–39.

Wolf, S. 2016. “Two Concepts of Rule Utilitarianism.” *Oxford Studies in Normative Ethics* 6: 123–144.

Woodard, C. 2019. *Taking Utilitarianism Seriously*. Oxford: Oxford University Press.

Notes:

([1]) See my 2010, 35–39 and my 2013, 723–724.

([2]) Most notably, Portmore (2011). See also Portmore, Chapter 1, and Hurley, Chapter 2, this volume.

([3]) Parfit (1984, 26); Feldman (1992, 182–185; 1997, 164–169); Portmore (2011); and Parfit (2017a, 395–406).

([4]) A fairly recent articulation of this idea can be found in de Lazari-Radek and Singer (2014, 312–313). Classic discussions are J. S. Mill’s references to secondary principles (Mill 1861, chap. 2); Sidgwick’s development of the idea that even act utilitarians should regularly think in terms of “common-sense morality” (Sidgwick 1907, Bk III, and Bk IV, chaps. III and IV); and R. M. Hare’s acknowledgment of the role of “intuitive-level” thinking (Hare 1981, chaps. 3, 8, 9). See also Woodard (2019, 195–200).

([5]) Brandt (1989, n. 22); Sidgwick (1907, 434–435); Pettit (1997, 97–102); Mason (1998, 386–393); Powers (2000, 239–254).

([6]) Cf. Wolf (2016, fn. 18).

([7]) Here I borrow words from Thomas Grey’s “Elegy Written in a Country Churchyard.”

([8]) The words are borrowed from John Rawls (1971, section 23). The idea that moral rules must be suitable to serve as *public* rules is implied in Baier’s (1958, 195f) and prominent in Gert’s (1998). I defended this idea in my 2010 (111–117) and 2016 (145–149).

([9]) See Slote (1984; 1985; 1989). For compelling arguments against satisficing consequentialism, see Mulgan (2001, 129–142). Cf. Chappell, Chapter 26, this volume.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([10]) A different view might be that attitudes (e.g., blame) are involuntary and thus beyond act-consequentialist assessment. But it seems to me that we do decide not only whether to express various reactive attitudes such as blame but also whether to have them. Be that as it may, the orthodox act-consequentialist position is that not only the expression of blame but also the having of this attitude can be assessed in act-consequentialist terms.

([11]) I first addressed this objection in my 1995. There I proposed that a rule-consequentialist agent's most basic moral motivation could be a concern for impartial defensibility rather than concern for the impartial good. See also my 2000 (chap. 4); 2007 (514–519); and 2008 (75–85).

([12]) For the methodology of seeking reflective equilibrium in ethics, see Rawls (1951; 1971, 19–21, 46–51). The point that rule consequentialism accords with our intuitions better than act consequentialism does was first made by Harrod (1936) and Urmson (1953). On the reflective equilibrium argument for rule consequentialism, see my 2000 (4–30). For a subtle and sophisticated development of rule consequentialism, see Mulgan (2006, chap. 5).

([13]) Compare Woodard (2019, chap. 5).

([14]) On the problem of different rates of acceptance, see Tobia (2013).

([15]) I am very grateful to Douglas Portmore for acute comments on an earlier draft of this paper. Many of the ideas canvassed in the revised version of the paper were ones he suggested that I discuss or at least admit.

Brad Hooker

Brad Hooker is Emeritus Professor at University of Reading and a Senior Research Fellow at Uehiro Centre for Practical Ethics at University of Oxford. He has published on a wide array of topics in ethics but is best known as the author of *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Understanding the Demandingness Objection

David Sobel

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.13

Abstract and Keywords

My aim in this article is to help us understand and assess the Demandingness Objection to consequentialism. I first try to motivate the Objection. Then I consider traditional replies that consequentialists have offered in an attempt to undermine the force of the Objection. Next I argue that for the Objection to be successful, it must explain which costs are deemed especially demanding and which costs are not, and why morality should be thought to prioritize the former. I show that the Objection cannot function as a persuasive critique of Consequentialism without prioritizing some costs over others. Finally, I consider reasons to doubt that the Objection can successfully meet this challenge.

Keywords: Demandingness Objection, consequentialism, well-being, ethical theory, morality, normative ethics

HARRIET Tubman made between thirteen and nineteen journeys back and forth from the free North to the Eastern Shore of Maryland to rescue between seventy and three hundred people who were enslaved there.[1] Every trip put her life in grave danger. Tubman had sustained a serious head injury as a child, when she herself was enslaved, which caused her to develop livelong attacks of narcolepsy and seizures. This condition obviously made such treacherous journeys significantly more dangerous.

She continued this work even after the Fugitive Slave Laws forced her to travel with the escapees all the way into Canada to avoid the threat of extradition. She would travel only in winter as fewer people were outside and detection less likely, yet the winter weather could be a great hardship for extended outdoor travel into Canada. There were rewards posted for her capture and her description was included, heightening the risk to her of detection and being re-enslaved. And of course her former enslavement, and having her head bashed in by the person who enslaved her, had vividly exposed her to just how awful this would be. And she well knew that the treatment she would receive if recaptured, after having escaped and famously helped others escape, would be dramatically worse. But of course this also gave her a more vivid sense of what she was rescuing people from.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

To give one a sense of proximity of danger that she regularly faced, in one instance she was attempting to rescue people in the same town that she herself had formerly been enslaved in. She saw the person who had formerly enslaved her and only managed to (p. 222) escape undetected because she quickly hit upon the plan of pretending that some chickens she was carrying had gotten loose, giving her the opportunity to run in the opposite direction from him, purportedly in chase after the chickens. Sometimes Tubman and the people she was helping escape would have to lay low for an extended period when bounty hunters had picked up their trail, and then lack of food would become a serious issue.

Later Tubman would work, without assurance of payment, for the Union army fighting for the end of slavery as a spy, scout, and guerrilla soldier. She would regularly successfully slip into territory controlled by the South and get information from enslaved blacks to take back to the Union army. She also worked for the Union army as a nurse caring for those with highly infectious smallpox. Later she would regularly put up relatives and homeless people in her own home. Toward the end of her life she used what resources she had to start up a home for the indigent. She died nearly penniless in that facility.

Tubman devoted her life to helping those in great need even when it involved extreme hardship, unthinkable danger to herself, and many others were much better positioned to aid. Her life shames us into an acceptance of how very much misery and oppression one can overcome if one is fearless, resolute, smart, resourceful, and compassionate, even if one has few other tools for such a task. To those of us who cower in fear at faculty meetings, unwilling to stick up for someone unjustly maligned, or who are unwilling to part with our creature comforts to prevent children from dying of dehydration, deciding to devote oneself to others to the degree Tubman managed is perhaps literally unimaginable.

1. Getting the Feel of the Demandingness Objection

Yet she might have done more. Even a life like Tubman’s, which is chosen because it exemplifies unparalleled and selfless service to bending what causal levers were available to her to alleviating misery and suffering, could have been lived in such a way to create even more goodness in the world. Imagine the complaint that Tubman did not adequately respond to the fact that other people’s well-being matters as much as her own. Imagine someone earnestly claiming that Tubman gave herself and her own comfort too much weight during her life to count as morally acceptable. She might, for example, have chosen to rescue people it was easier to reach rather than focus on areas where she had family and, in so doing, rescued a few more people. She probably could have squeezed in at least one more trip south to help a few more people escape enslavement.

Plainly such a complaint against Tubman would be outrageous. If Tubman does not measure up to some purported moral standard on the grounds that even she was not (p. 223) giving enough of her time and energy, if even she counts as immorally selfish by its lights,

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

then that moral standard is not suitable for human beings. Consequentialism, because it tends to insist that acts either maximize goodness or are wrong, therefore, is not a suitable understanding of morality for human beings. One might well maintain that condemning Tubman as immorally selfish is merely the pinnacle of the mountain of absurdity involved in demanding that agents maximize the good if they are to count as morally acceptable.

Let's call that broad concern the Demandingness Objection to consequentialism. Other moral theories might be thought to be too demanding as well, but this concern is thought to apply paradigmatically against consequentialism, and we will consider it only in this context. Of course, friends of the Objection maintain that we can promote the good radically less vigorously than Tubman and still launch a successful Demandingness Objection against consequentialism in defense of the moral acceptability of our wildly-short-of-Tubman-levels-of-promotion-of-the-good lifestyle. Friends of the Objection maintain that it can be used to defend the moral acceptability of the lives of ordinary folk, not just moral superheroes.

The case of Tubman, and the outrageous complaint that she did not do enough to count as living a morally acceptable life, is meant to put us in a position to feel the intuitions that animate the Demandingness Objection. Now let's try to put the intuition into words, that is, formulate an initial general and broad commonsensical handle on the complaint the Objection offers. I think the intuitive thought is something like this: Morality should not take over our lives, at least in most circumstances and certainly not in circumstances such as most of us live in today, but rather will be fully compatible with a wide range of self-directed lives that involve significant devotion to friends, family, and/or projects other than serving morality or the interests of others. Morality, except in very rare circumstances, leaves us plenty of free space to construct a life according to our own lights. The true morality will not dictate the most central and pervasive aspects of our lives. This is why Tubman was certainly not required to do anything like the work she did for others, even if that work was vitally important in making the world contain a much greater amount of goodness. Moral views that claim the only way to be morally acceptable is to fundamentally and thoroughly shape our lives to conform to its demands are mistaken. The true morality, the Objection insists, will reflect the truth that there are a wide range of permissible ways of life that provide the opportunity for people to fashion a life suited to themselves.

This broad, and admittedly in some ways still vague, concern could be fleshed out in various ways. It is so far intended merely as a common creed that different versions of the Objection will make more precise. Demandingness might be measured in different ways. The Objection might be best understood as claiming that a moral regime is unrelentingly demanding to her—it never provides contexts, such as a weekend, where it relaxes its demands. Or it might be that a regime excessively constrains the complier's options, even if they are not too awful with respect to opportunities for her well-being. Or it might be that

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

the expected costs of a purported moral regime are too high. Such differences can safely be ignored for our purposes here.

(p. 224) We will, however, spend some time later in this article thinking about another difference in understanding the Objection. The problematic issue that the Objection points to might be claimed to be that compliance with an understanding of morality would be too costly to the complying agent in terms of her own well-being. Or it might be that what this moral regime requires is too costly, even if it does not demand all those costs voluntarily of the agent but rather imposes some of them on a person qua patient rather than agent of morality. To see the difference between these options, consider the difference between a moral regime requiring that one freely give over one's kidney oneself as opposed to requiring that other people take your kidney from you. Does the owner of the kidney have an equal demandingness objection against both regimes, or is it more powerful against one or the other?

One issue at stake here is whether we want to mark an important difference relevant to understanding the Objection between a "demand" of a moral theory and a more generic "cost" to an agent that results from that theory. One might, for example, say that when a moral theory requires that others take one's kidney, that moral theory is costly to you but does not demand anything of you since its requirements do not address you qua agent. On this understanding, only costs a moral regime requires of one qua agent count as demands of that regime. The word "demand" plausibly does seem to have such a connotation. However, for the Objection to lean on linguistic aspects of the word "demand" in pressing this distinction threatens to rob the Objection of much of its force unless some moral significance can be given to the distinction between what a theory demands in this sense and what costs it imposes on agents more generally. If the Objection narrows the understanding of "demands" in this way without providing a rationale for thinking such costs are morally prioritized, we should lose interest in the resulting understanding of "demands" in assessing moral theories. We will return later to consider one reason to accord this distinction moral significance—the insistence that morality must provide an agent with strong reasons to comply. This thought might rationalize looking especially at costs imposed on agents when they are being required to act by a moral theory since, on this understanding, such costs cannot in principle get so steep as to make it the case that the agent does not have strong or decisive reasons to comply.

2. Traditional Lines of Defense from the Objection

Defenders of consequentialism from the Objection have, sensibly, not been keen to denounce the immoral selfishness and sloth in the service of the good of Tubman. Instead, they have developed a Swiss Army knife of replies to the Objection.[2] Here is (p. 225) a quick, and surely incomplete, list of what seem to me the most important such replies: First, consequentialism is an account of the moral truth-maker, not a decision procedure. Thus morality does not require that we keep our eye on the task of promoting aggregate

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

utility. Because of this, it is sometimes suggested, lives that are not obsessed with or self-consciously focused on aiding other people can be morally acceptable.[3] Second, most who make no or little time for friends and loved ones, because such time and energy would siphon off resources that could create more goodness elsewhere, tend to burn out in the long term and become ineffective in promoting the good.[4] Thus the wise promoters of goodness, who recognize human weaknesses and their own susceptibility to it, will not deprive themselves of such support. And this will require devoting time and energy into sustaining friendships so that they can sustain you. Third, various alternative forms of consequentialism, be it rule, motive, cooperative scheme or whatever, will reveal a less demanding, and partly for that reason, more compelling form of the view.[5] Fourth, morality only requires that we produce enough of the aggregate well-being, or goodness, we might create, perhaps 70 percent. Satisficing rather than maximizing versions of consequentialism will be more tolerably demanding.

Fifth, scalar forms of consequentialism deny that there is a cut-off line of permissibility, but rather only better and worse moral options. If we accept this, we will see that consequentialism makes no "demands" at all and thus could not be too demanding.[6] Sixth, some consequentialists might join some virtue ethicists in maintaining that lives that sacrifice moral excellence cannot do so to the genuine prudential advantage of the agent making such decisions. If it really were true that the best life for one's own well-being was the morally best life, then seemingly demanding moral theories, if otherwise true, are not asking for a sacrifice of the agent's well-being. In a sense, the thought is, it is prudence that is demanding, not morality.[7] And obviously prudence is not demanding (p. 226) in the coin of well-being, which might well be thought that most salient coin of demandingness.[8] Seventh, others maintain that the best version of consequentialist morality only requires that we do our fair share to make the world contain as much good as possible. If so, morality's demands would look much less demanding and much more like we are used to.[9] Eighth, even if Tubman's life was not, at all points, morally acceptable, saying or even thinking as much is unjustifiable by the true moral standard. Indeed, the only attitude permissible toward Tubman is praise and admiration. The true consequentialist will measure the moral appropriateness of the reactive attitudes and their expression by considering if such expressions themselves maximize goodness. Given how little more Tubman might have done, and how much more many others might do, and how unlikely we are to persuade many to do anything remotely as onerous as Tubman, blaming Tubman will turn people off the project of adhering to morality and cost more goodness than it produces. It would be expedient in terms of promoting the good to lavish praise on the efforts of Tubman and immoral to denounce her as immoral.[10]

Ninth, still others suggest that concerning ourselves with only our own sphere of friends and family will, most of the time and in ordinary circumstances, in fact be the best way to produce the most overall utility. Because we are closer to these causal levers and more knowledgeable about what in fact would benefit people we are close to, focusing on those one knows well will often be the best way to produce the most good.[11] Tenth, some distinguish objective and subjective compliance with morality's demands. Subjective compliance is secured when, given one's information, one does the best one reasonably could by

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

way of making the world a better place. If one does that, yet fails to in fact maximize goodness, one is not thereby blameworthy as blameworthiness is tied to subjective failings. Eleventh, some point out that in a world without slavery and with a much more egalitarian distribution of wealth, the demands of consequentialism would be greatly diminished and the occasions in which one's doting on oneself and one's inner circle would maximize goodness would be greatly expanded. The great demands of consequentialism are a result of the great evils and inequality in the actual world, not intrinsic to the view. Such people will encourage us to blame the state of the world, or (p. 227) those who have created the modern horrors, not consequentialism, for the demands of morality in this world.

Twelfth, others point out that many cases in which consequentialism seems very demanding are cases where the vast majority of people in a position to aid are not doing their fair share to promote the good. If others would do their fair share, the burden on each person would be greatly reduced. We should not blame consequentialism for the great burdens that fall on the few who will listen to morality. The blame is properly directed at those not doing their share. Thirteenth, we must distinguish between something being morally required of X and some Y being entitled to force X to do that thing. Consequentialists can allow that moral obligations can exist even when no one is entitled to enforce them. In such cases one might say the burden is too demanding to be enforceable, even if it is still a moral demand on one. This would be an attempt to reinterpret the demandingness worry as a political rather than moral worry.

Broadly speaking, and with some exceptions, these responses primarily aim only to show that the correct understanding of consequentialism is not so demanding that we should denounce Tubman as immorally self-centered. The replies try to make room to think that consequentialism is more demanding than mainstream understandings of morality, but not so extremely demanding as to be obviously unacceptable. Broadly, and with exceptions, the typical consequentialist reaction to the Objection is to accept that there is some truth to the charge of the Objection but to try to mitigate the damage to the view with the aforementioned maneuvers.

Without meaning to cast doubt on the truth of any of these replies, or to endorse any of them, I would like to confront the Demandingness Objection in a different way. I want to query how we should understand what it is to be demanding in the first place and why it should be thought that morality is not demanding in that sense. We need to understand the Objection better before we can see how the consequentialist might best attempt to respond.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

3. The Ambitions of the Demandingness Objection

But before getting to that, I need to say a few words about how I will be understanding the ambitions of the Objection, as that plays a crucial role in my concerns about its success. An unambitious understanding of the Objection would have it merely claim that consequentialism claims some things are morally required when our intuitions maintain that such options are merely permissible and not required. And on this basis the unambitious interpretation would conclude only that this is some reason to doubt that consequentialism is correct in those claims about what is required. The unambitious understanding is fully compatible with the thought that, for example, our demandingness intuitions are sensitive to the truth that there are nonconsequentialist rights people (p. 228) have that prevent some options that consequentialism thinks are required from being required or that there is a morally important difference between causing and allowing that consequentialism misses and that our intuitions of demandingness are picking up on cases where consequentialism is inappropriately holding us equally accountable for things we allow as for things we cause.

This unambitious understanding of the Objection is compatible with the idea that our demandingness intuitions merely signal the existence of some other, fundamental problem with consequentialism and that the experience of excessive demandingness is downstream from the genuine and fundamental problems for the view. On this view, the Demandingness Intuitions are not insisted to be more than merely an epistemic guide to finding the real problem with consequentialism, which is more fundamental than, and explains, our intuitions about excessive demandingness.

A more ambitious understanding of the Objection would have it maintain that a crucial problem with consequentialism is that it fails to accept that the true morality is not excessively demanding. This understanding maintains that we can understand the level of demandingness of a moral theory significantly independently of a commitment to a particular moral outlook and assess the adequacy of that moral outlook on the basis of how demanding it is. On this understanding, our demandingness intuitions don’t merely signal or track some possibly independent fundamental problem with consequentialism, but rather the Objection is a self-standing and independent objection to consequentialism. Hereafter I will only be addressing the more ambitious version of the Objection. It is this understanding, I believe, that most who pressed the Objection and most who defended consequentialism from it have had in mind.

4. Toward Understanding the Demandingness Objection

In paradigmatic types of cases in which people feel the pull of the Demandingness Objection, there is a purported moral requirement that X aid Y that is felt to be excessively de-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

manding on X. For concreteness, let's imagine a situation in which Sally nonculpably needs Harry's kidney. Harry can live without this kidney, but Sally cannot. Still, most who champion the Objection would claim that this purported requirement on Harry to give over the kidney is excessively demanding on Harry and so not a genuine moral requirement.[12] Let's suppose this requirement would count as excessively demanding in the sense that the Objection is pressing. With that supposition in place, we can look at features of the case to try to help us more precisely interpret what it means to say that something is too demanding.

(p. 229) Obviously, in our scenario, the costs to Sally of Harry's failure to give over the kidney are larger than the costs to Harry of doing so. But no one would for a second maintain that Sally has a Demandingness Objection against a moral regime that permits Harry to not hand over the kidney. When we are in the grips of the Demandingness Objection intuitions, we are focused on costs to the aider and not on the larger costs that would befall the potential recipient of aid if she were not aided. Only because this is how we are thinking of what is and is not demanding does it appear that consequentialism is more demanding than some nonconsequentialist moral regime in cases of this sort. If the costs of what a moral theory permits to befall the unaided were thought just as demanding as costs a moral theory imposes on someone by requiring him to aid, then, in our examples, the Demandingness Objections would be leveled more forcefully against the nonconsequentialist moral views that do not require that the aid be provided.

Thus, a crucial part of clarifying the Objection's notion of what is and is not demanding must involve the Objection prioritizing some costs that would result from a moral regime over other such costs. I want to pause and underline this point. I think the lack of attention to this point has damaged much of the discussion of the Objection. Any potential vindication of the Objection must maintain that not all costs of a purported moral regime are on a par in their demandingness. Some types of costs that result from a purported moral regime must be maintained to be more demanding than others even when costs of other types are larger. If you take only one lesson from this paper, let it be this: friends of the Objection must specify which types of costs are properly prioritized by the Objection. And they must then address the question of why morality, properly understood, should prioritize such costs.

So any potential vindication of the Objection must address the question of which types of cost are especially demanding. The claim that all types of costs count as equally demanding will not get the result friends of the Objection are looking for in paradigmatic cases where people feel the force of the Objection. Further, that no one presses the Demandingness Objection on Sally's behalf but only on Harry's behalf, despite the costs to the former being larger, is an important clue in how to understand what is, and is not, seen as especially demanding by the Objection.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

5. Potential Problems for Understandings of Which Costs Are Especially Demanding

Whatever costs the Objection prioritizes as especially demanding face a series of questions which have the potential to diminish the Objection from playing the role it is typically thought to be able to play. Here I will just mention a few such questions. Later we will press these questions on various interpretations of the costs prioritized by the Objection. The interpretations offered in this section are not mentioned because they (p. 230) are especially plausible interpretations of which costs are prioritized by the Objection. Rather, they are selected because they especially clearly illustrate a potential problem that could arise for other, more plausible, understanding of which costs are prioritized by the Objection.

Some possible understandings of the costs prioritized by the Objection would reveal the Objection to be downstream of the important criticisms of, or fundamental break with, consequentialism. For example, suppose the costs prioritized by the Objection were costs involved in doing what one is morally permitted but not required or forbidden to do. That is, we first figure out what options morality requires one to do and which options are not required. Then we assign higher demandingness to costs a person bears in doing things she is permitted but not required or forbidden to do. So if Harry is permitted but not required to give over his kidney, then costs associated with giving it over would count as especially demanding. If Sue is required to return the money she borrowed, the costs to her of doing so are not prioritized by the Objection. If Pei Lin is not permitted to steal Yousef's car, the costs to her of not doing so are not prioritized, on this picture, as demanding.

This understanding of the costs that are especially demanding would obviously reveal the Objection to not be the fundamental critique of consequentialism. For on this understanding of what makes costs especially demanding, the Objection is forced to presuppose that consequentialism is incorrect as a premise in its account of which costs are and are not demanding. The Objection, in this case, would be riding the coat tails of a previous break with consequentialism, not itself providing the rationale to break with it. Call this the Redundancy Worry. In the earlier case it is transparent that the Redundancy Worry applies. However, in the case of some more plausible interpretations of which costs are prioritized as especially demanding, there can be nontransparent Redundancy Worries.

Second, suppose the costs that are prioritized are just any costs an agent has the power to avoid paying. That is, if morality purportedly requires Superman to wait in line but he can zoom to the front and no one can stop him or impose costs on him for doing so, then such costs on Superman count as especially demanding. On this interpretation of what is demanding, it might seem that the Objection is here engaged in a kind of negotiation with people who especially have the power to resist the call of morality without harm to themselves. If it were thought that people care somewhat to be moral and to be seen to be moral but tend to be reluctant to significantly sacrifice their own well-being or what they

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

care about, it might be wise for a moral regime to diminish what it asks of such people who can relatively costlessly revolt against morality. The thought might be that if we ask for all that morality in fact requires, we will get less out of people who are significantly self-interested and in a position to resist costs imposed by morality. Better, from the moral point of view, to purport that morality requires less of people in such a situation so that we get what we can out of them. And of course, on this interpretation, in some contexts some of us are relevantly like Superman and in other contexts we are not. That is, in some contexts some of our actions are difficult to detect or performed before people poorly positioned to effectively object and other of our actions are not.

(p. 231) This understanding of what is demanding would reveal the Objection to not be in genuine conflict with consequentialism. Consequentialism is an account of the truth-maker of moral claims and only indirectly takes a stand on what moral requirements it is wise to publicly announce.

Alternatively, and depending on our own theory of reasons for action, this understanding of the costs that are especially demanding might just reflect which costs we think an agent has reason to pay, rather than which costs morality requires. If we reject a moral rationalism which claims that people always have most reason to do what morality requires, there can be cases where morality requires X but the agent lacks most reason to X. The Demandingness Objection might register the thought that the agent lacks a powerful reason to comply, not that morality does not ask her to comply. So understood, it would seem the Objection would not tell against consequentialism as a moral theory. Call these two thoughts the Not an Objection to Consequentialism Worry.

Third, suppose the costs prioritized by the Objection were costs just to people who are well enough off that they are frequently in a position to aid others. That is, any cost to any such person is seen as especially demanding. In this case we would be right to worry that the Objection is merely protecting the privileges of the privileged and unconsciously responding to the morally inappropriate sense that somehow high-status people are more morally important than other people. In this case we would likely psychologize the intuition, perhaps by suggesting that either the people whose intuitions were consulted were those of people already so privileged, expected to become so, or sought to curry favor with (or avoid the wrath of) the privileged.[13] Alternatively, we might wonder if society has managed to deceive even nonprivileged people to unconsciously react as if the privileged mattered more by telling stories that focus on the heroism and self-made nature of the well-to-do and the debased nature and deserved status of the poorly-off. With this interpretation of which costs count as especially demanding, we ought not find the Objection so understood to be morally compelling. Call this the Affluenza Is Contagious Worry.

6. Which Costs Are Especially Demanding?

I now want to consider a few candidates for the sort of costs best prioritized as especially demanding by the Objection. Obviously I will not be able to consider every possible understanding of which costs are especially demanding and how prioritizing such costs

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

fares with respect to all the types of challenges I mentioned earlier. Rather, I will consider what seem to me the most plausible such understandings I can think of and (p. 232) consider the most pressing problems for each stemming from the type of worries raised in the previous section.

We will consider two possible understandings of which costs are prioritized by the Objection. Let's start by considering the view that the costs that are especially demanding are what I will call active compliance costs. These are costs to the actor in her voluntarily doing what is purported to be a moral requirement.[14] This proposal would explain our sense that the costs of requiring a person to give over her kidney generate a powerful demanding objection against such a requirement but costs to the person who would die if the kidney is not given over do not similarly generate as powerful of a demandingness objection, despite the costs to the latter being higher. More generally, this understanding of what costs are especially demanding would explain why our demandingness intuitions are much more sensitive to costs to the potential aider than to the person who would suffer greater costs if left unaided.

There are a variety of concerns to have about this proposal. I want first to list those concerns and then offer what I think is the best response to such worries. So, to begin, let's consider the worries about the proposal. First, it is odd to ignore benefits that flow to the person qua patient of a moral regime. If an agent is better off overall under a moral regime, but pays higher compliance costs compared to a rival moral regime, it is odd to think of that moral regime as more demanding on that agent such that the agent has a stronger complaint against it than against a rival moral regime which would leave her, all in, worse off. It is odd to say that a moral system is too demanding on one when one would be better off under that moral system than under any rival moral system. Second, we must wonder what to think of costs imposed on one by a moral regime qua patient. Can I launch a Demandingness Objection against a moral regime that requires others to take my kidney? If not, is there some other complaint to lodge, such as a failure to respect my rights, which should not be thought of as part of the Objection? Making this a separate complaint threatens to make the Objection seem rather narrow. Third, prioritizing active compliance costs responds to the thought that morality ought not take over our lives. However, it suggests that it is more morally important that morality not take over our lives than that disease and famine not do so. This priority seems morally suspect and, contrary to what I would expect morality's priority to be, tailored to suit the powerful rather than the needy.

(p. 233) Fourth, one rationale for prioritizing active compliance costs is because these are costs a person must voluntarily impose upon herself. And if such costs are too high, people may well choose to not impose them on themselves. Thus we may do better in terms of the bottom line if we lower the costs of people feeling like they are adequately responsive to moral concerns. This is a strategic move, not one generated by intrinsic moral considerations. This should remind us of one variant of "Not an Objection to Consequentialism" earlier. Fifth, it is more obvious that an agent's reasons are especially responsive to costs to her that she can choose to avoid than that morality is especially responsive to

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

such costs. It is widely accepted that one has special reasons to be partial to oneself. Indeed, many maintain that one can have sufficient reasons of partiality to rationalize doing so over the requirements of morality. What is much less clear is that morality itself provides some ground to prioritize costs to the deliberating agent over larger costs to others. Again we might think what we are responding to when we feel the Demandingness Intuitions is that the agent does not have a strong reason to comply with morality, rather than thinking that morality does not really require it. This should remind us of the other variant of "Not an Objection to Consequentialism" earlier.

Sixth, the simplest way to overcome the Objection on this understanding is simply to impose fewer active compliance costs and greater passive compliance costs. That is, stop asking people to give over their own kidney voluntarily and start asking people to forcibly take another person's kidney. If a moral regime stops asking us each to voluntarily turn over our kidneys and instead asks us each to forcibly take the other person's kidney, that moral regime will, on this understanding of demandingness, greatly reduce its demandingness. But the latter, purportedly less demanding, view seems plainly a less good moral view. Such a view seems all in worse and to not have an important virtue compared to the version that imposes such requirements on us as agents. It is a bad sign if we significantly reduce the demandingness of a moral regime on an understanding of demandingness, keep it otherwise as similar as possible, and thereby make the moral regime worse. That is a sign that this understanding of what costs are demanding are not getting at a very morally significant complaint against a moral view.

Seventh, if a competent person voluntarily and under no duress makes a bunch of contractual promises to aid people, this surely diminishes her ability to complain about the demandingness of a moral theory that requires her to live up to those promises. However, it does not seem the proposal under consideration can make room for this thought. The active compliance costs of aiding others, even after you have freely promised to do so, is just as high as it would have been had one not promised. Thus it seems on this proposal the size of the demandingness complaint a person can make against being required to aid is implausibly unaffected by her free promise to do so. Surely one cannot run around making lots of uncoerced promises to help people and then legitimately complain about the demandingness of a morality that would require one to keep one's word.

Now let's consider the best reply to some, but not all, of the aforementioned concerns. Fiona Woollard has helpfully pointed out that the focus on avoidable costs to the moral agent of complying with purported moral demands could be understood to reflect the (p. 234) thought that "morality should be such that it is generally reasonable to expect an agent to choose to conform to it. The Demandingness Objection charges that the target theories do not meet this condition."[15] If some rationale for a kind of moral rationalism that forges a strong connection between an agent's reasons and what morality asks of her could be provided, several of the aforementioned concerns might be significantly allayed.[16] Let's call this the Rationalism Rejoinder.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

I have two main responses to the Rationalism Rejoinder. First, it is crucial to note that this response needs to rely on more than just there being such a strong connection between an agent's reasons and what morality asks. It also needs to assume a particular picture of what sorts of things an agent has reason to do. If a pure subjective view about reasons were correct, and agents only have reason to do what furthers their contingent ends, the strong connection between reasons and morality will look much less plausible.[17] On the other hand, if we have powerful reasons to promote what is objectively good, our reasons might be quite demanding themselves. If subjective accounts are false, it may be that one has powerful reason to sacrifice significantly for others, at least when others are in desperate need and one can do so without sacrificing anything of comparable importance. Alternatively, one might think that whatever the content of the true morality, we have very strong reasons to obey it. This would make our reasons bend to what morality independently requires, blocking the claim that morality must bend to accommodate our reasons.

More broadly, the persuasiveness of this sort of defense of the Objection requires a kind of confidence in a theory of reasons which is partial enough to vindicate the Objection but not so selfish that something deserving the name morality cannot accommodate itself to it. The move under consideration here requires not only that morality typically provides such reasons, but also that an agent's reasons are especially responsive to costs to herself, even when countervailing moral considerations are in play. Some forms of rationalism, or the insistence of a tight connection between one's reasons and what morality requires, will, as it were, elevate one's reasons to morality. Others will cut morality down to what one has reason to do. For the response here to work it needs to have something closer to the latter structure. The response here must assume a rather partial picture of what one has reason to do together with a strong connection between (p. 235) one's reasons and the requirements of morality. I think these two components are less tempting in conjunction than either is separately.

My second reply to the Rationalism Rejoinder is that it still seems vulnerable to the Redundancy Worry. Consider a case where a nonconsequentialist moral regime prohibits a person from taking another person's kidney without that other person's consent, even though the agent will die if she does not take it. As we are currently measuring demandingness, such an agent has a very powerful Demandingness Objection against this moral regime based in the fact that her complying with the regime would be extremely costly to her. Presumably, however, the friend of the Rationalism Rejoinder still wants to get the result that despite such a strong Demandingness Objection against this requirement, the agent is not all-in permitted to take the kidney. The only way I see to get this result is to somehow maintain that the force of the reasons morality provides to not take the kidney outweigh the force of the Objection in such cases. But to get this result, together with the result that agents are permitted to not give their kidney in such situations, it must be presupposed that morality provides stronger reasons against causing the loss of a kidney than it provides against allowing a person to die unaided from a lack of a kidney—that is,

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

to presuppose a strong version of the causing/allowing distinction. Thus the resurfacing of the Redundancy Worry.

A second way we could think of the costs that the Objection might champion as especially demanding would be to prioritize costs that are required by a moral regime whether those costs fall on people in their role of agent or as patient of the regime. On this picture costs of being required to, qua agent, voluntarily give over one's kidney and costs to Y when a moral regime requires X to take Y's kidney without Y's consent would both count as especially demanding. But if both such costs are prioritized, what costs are demoted in significance on the demandingness scale? For some costs to be prioritized, other costs must be diminished in importance.

This picture would downgrade the moral significance of costs permitted but not required by a moral regime.[18] So if a moral regime required someone to lose a kidney, the costs associated with that loss would count as especially demanding but if it only permitted someone to die from a lack of a kidney, rather than requiring that action be taken to prevent that result, the cost of that lack would not be prioritized by the Objection. In this way the Objection could again prioritize costs to the aider or the person who others are required to use as a mean in aiding, but it would downplay the costs of those who will go unaided if the moral theory does not require the aid to be given. This, at first blush, has the right shape to vindicate the traditional focus on costs to those required to aid and relative downplaying of costs that will result to those who the theory permits to go unaided.

And this suggestion fits well with the name of the Objection, the "Demandingness" Objection. Costs permitted by a moral regime but not required do not count as "demanded" (p. 236) by that moral theory. And so this is a second reason to think this proposal a plausible interpretation of the costs best understood as prioritized by the Objection. Further, this proposal is less plausibly interpreted as merely a pragmatic bargain with the agent or responding to that agent's reasons rather than a moral assessment. The suspicion that we are talking about the agent's reasons rather than morality arises especially in contexts where we are fixated on costs to the acting agent.

So, having motivated this interpretation, let's begin to assess it. The first and most important thing to say here is that this distinction between costs required and costs permitted looks a lot like the causing/allowing distinction adjusted so as to assess moral theories rather than agents. Moral regimes don't cause stuff. But they can require stuff. And in a morally ideal world such requirements would affect people's behavior. Moral regimes also tolerate stuff by not requiring what would be needed to prevent it. The current proposal supposes this distinction is morally significant. One way of arriving at such a view might be to say things like: if a moral regime permits X to occur, we should not think the regime necessarily is in favor of X or should be held responsible for X. But if it requires that X occur, then it does seem that in some sense the regime is in favor of X and, should people respond to the requirement appropriately, could legitimately be thought responsible in part for X. And it might seem more appropriate to count such costs against a regime when that regime is for those costs being paid and partly responsible for them being paid.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Such thoughts may be sensible, I will not here argue that they are not, but they are the same sort of thoughts that lead people to mark a moral distinction between costs caused by X and costs allowed by X. After all, in such cases, we might say that in cases where X merely allows stuff they need not be for such costs or responsible for them. That is to say, the same sort of thoughts that might lead one to mark a distinction between costs caused versus allowed by agents are the thoughts that might lead one to mark a distinction between costs required and permitted by a moral regime, and to morally prioritize the former over the latter.

Thus I think that the move of prioritizing costs required by a moral theory and downplaying the costs allowed presupposes a fundamental and familiar sort of break with consequentialism in how it measures demandingness. And the distinction between causing and allowing or requiring and permitting is, I submit, plainly an independent break with consequentialism. Thus the fundamental break with consequentialism is independent of and downstream from worries about demandingness. This is, obviously, just an instance of the broader Redundancy Worry we saw earlier.

Obviously some ways we might reject consequentialism, such as by marking a morally crucial distinction between causing and allowing and maintaining that we are less morally accountable for what we allow than for what we cause, will imply that any view like consequentialism, which maintains that we are just as accountable for what we allow, is therefore too demanding with respect to what we may allow. But that worry about demandingness will just be a trivial entailment of a prior and independent ground for rejecting the view.

(p. 237) 7. Conclusion

A persuasive version of the Demandingness Objection must successfully perform three crucial tasks. First, it must tell us which costs it is prioritizing as especially demanding. Only if the Objection prioritizes some costs over other larger costs can it hope to vindicate the claim that consequentialist theories are especially demanding, as we saw in the kidney example. Second, it must persuade us that such costs are morally prioritized. And third it must not presuppose prior and independent breaks from consequentialism in measuring the demandingness of an ethical theory. Until proponents of the Objection have offered us a persuasive understanding of the Objection that persuasively addresses these three topics, I maintain that we should reject consequentialism independently of the Objection or not at all. Until we are offered such an understanding of the Objection, the Objection itself does not provide a good reason to reject consequentialism. In my view, full recognition of these burdens in successfully defending the Objection have only relatively recently been acknowledged and responded to. We may be in the infancy of our understanding of the Objection.[19]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Notes:

([1]) I have previously written on this topic in my "The Impotence of the Demandingness Objection," *Philosophers' Imprint* (Sept. 2007) (also collected in my *From Valuing to Value* [Oxford University Press, 2017]). I continue to accept what I wrote back then. A few points from the earlier work are again insisted upon here. However this article is not intended as an update of, or to replace, that earlier paper.

([2]) Not all of these replies make a direct case against the Objection. They do not all make a case that morality, properly understood, may take over our lives in the way the Objection denies. Rather, some of them try to persuade us that consequentialism can find various other ways successfully to give vent to our most powerful intuitions in the neighborhood of the Demandingness Objection in ways that relieve the degree of conflict with common sense.

([3]) While many have defended the claim that consequentialism is best understood as a truth-maker rather than a decision procedure, I find the discussion in Parfit's *Reasons and Persons* (Oxford: Oxford University Press, 1984), especially forceful.

([4]) Peter Railton, "Alienation, Consequentialism, and the Demands of Morality," *Philosophy and Public Affairs* 13, no. 2 (Spring 1984), forcefully defended such thoughts.

([5]) Brad Hooker, *Ideal Code, Real World* (Oxford: Clarendon Press, 2000), has suggested that part of the rationale for looking to rule views is to avoid the Demandingness Objection. However, he is not committed to the idea that the resulting view remains an instance of consequentialism. In one sense it is not really a defense of consequentialism from the Demandingness Objection to urge us to reject consequentialism and move to a different view. It does, however, show that one accords the objection real force.

([6]) Alistair Norcross argues for such a view in his *Morality by Degrees: Reasons without Demands* (forthcoming).

([7]) It would be interesting to consider the force of a Demandingness Objection against a theory of prudence. Could one similarly complain that a purported prudential regime took over one's life and provided insufficient freedom to fashion a life of one's own choosing? Perhaps alienation concerns about some objectivist theories of well-being could, within the bounds of tolerable revision, be understood along these lines.

([8]) Rosalind Hursthouse, *On Virtue Ethics* (Oxford University Press, 1999), for example, has argued that a virtuous life is always prudentially recommended, given one's epistemic situation, even if in some rare and unlikely cases such virtuous lives can turn out to not be good for the person who lives them. I take issue with such a view in David Copp and David Sobel, "Morality and Virtue," *Ethics* 114 (April 2004): 514–554.

([9]) Liam Murphy, *Moral Demands in Non-Ideal Theory* (Oxford University Press, 2000), argues for such a position. It is not entirely clear to me if he thinks of the resulting view as still an instance of consequentialism.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([10]) Peter Singer and Katarzyna de Lazari-Radek, *The Point of View of the Universe: Sidgwick and Contemporary Ethics* (Oxford University Press, 2015), chapter 11, have argued in this direction.

([11]) Mill unconvincingly suggests this reply in *Utilitarianism* (Indianapolis: Hackett, 2001). Given the ease with which resources and information today can move about the world and the wildly nonegalitarian distribution of resources together with desperate need that it takes little insider information to understand how to fix, this suggestion seems incorrect about the world we in fact live in even if it could be correct about some possible worlds, including some worlds that were actually in the past.

([12]) I used this example in my "The Impotence of the Demandingness Objection."

([13]) John Harris, "The Survival Lottery," *Philosophy* 50 (1975): 81–87.

([14]) Understanding what we are trying to measure here is not trivial. Seemingly if there are costs involved in not doing what is purportedly morally required, such as costs of ill will from one's neighbors, incarceration, or maybe even guilt, this should be thought to reduce the size of the compliance costs of complying. That is, the best understanding of the size of such costs seems to involve a comparison between what would happen if one complied and what would happen if one did not. Thus larger costs involved in failing to comply would reduce the relevant compliance costs. If that is right, perversely the compliance costs involved of living up to morality in a highly moralistic and surveillance-ridden society go down. If so, then the demands of consequentialism in a society that stands willing to detect and punish those who do not effectively further the good would be smaller. There would thus be different ways to reduce the demandingness of consequentialism to acceptable levels, only some of which changes what the theory requires of agents.

([15]) Fiona Woollard, "Dimensions of Demandingness," *Proceedings of the Aristotelian Society*, vol. cxvi, Part 1 (2016), 94. See also Brian McElwee, "What Is Demandingness?" in *The Limits of Moral Obligation*, edited by Marcel van Ackeren and Michael Kühler (New York: Routledge).

([16]) For my favorite type of argument for such moral rationalism, see Stephen Darwall, *The Second Person Standpoint* (Cambridge, MA: Harvard University Press, 2006), 98; and Doug Portmore, *Commonsense Consequentialism* (Oxford: Oxford University Press, 2011). I consider this line of thought in my "Subjectivism and Blame," in *Reasons to be Moral Revisited*, ed. Sam Black and Evan Tiffany, *The Canadian Journal of Philosophy*, Supplementary Volume 33 (2009): 149–170.

([17]) I have argued toward subjectivism about reasons in "The Case for Stance Dependent Reasons," forthcoming in *Journal of Ethics and Social Philosophy*; "Subjective Accounts of Reasons for Action," *Ethics* 111 (April 2011); and "Subjectivism and Idealization," *Ethics* 119 (January 2009). These, and other papers in this vein, are collected in my *From Valuing to Value* (Oxford University Press, 2017).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([18]) Liam Murphy explicitly endorses such a proposal. See his excellent general discussion of such issues in *Moral Demands in Non-Ideal Theory* (Oxford University Press, 2000). I reply at length to Murphy's argument for this position in my "The Impotence of the Demandingness Objection."

([19]) Doug Portmore offered me really helpful comments on an earlier version of this article.

David Sobel

David Sobel is Guttag Professor of Ethics and Political Philosophy at Syracuse University. He is the author of From Valuing to Value (Oxford University Press, 2017), founding coeditor of the Oxford Studies in Political Philosophy series, and coeditor of the blog PEA Soup. His primary research project focuses on the question of what makes things valuable. He is especially interested in whether, and to what extent, it is our attitudes toward things that make them valuable for us.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Consequentialism, the Separateness of Persons, and Aggregation

David O. Brink

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.14

Abstract and Keywords

This essay reconstructs and assesses claims that utilitarianism and, more generally, consequentialism have inadequate conceptions of distributive justice, because their aggregative character ignores the separateness of persons. On this view, the separateness of persons requires a fundamentally anti-aggregative conception of distributive justice. Even if this objection applies to some forms of utilitarianism, it won't apply to forms of consequentialism that recognize some conception of distributive justice as an important nonderivative good. Moreover, the separateness of persons poses, rather than resolves, questions about the role of aggregation within distributive justice. This essay explores the adequacy of some consequentialist answers to these questions and defends selective, rather than unrestricted, aggregation.

Keywords: aggregation, consequentialism, contractualism, distributive justice, egalitarianism, separateness of persons, sorites, transitivity, utilitarianism

UTILITARIANISM is a consequentialist moral theory that takes the good to be promoted—in one formulation, maximized—to be the general happiness or welfare. It takes everyone's interests into account by aggregating their interests, balancing benefits to some against harms to others, as necessary, so as to produce the best total outcome. This conception of impartiality is aggregative and permits interpersonal balancing of benefits and harms. But some critics of utilitarianism claim that whereas balancing goods and harms *within a life* might be acceptable, balancing goods and harms *across lives* is impermissible. This is because of the *separateness of persons*. Intrapersonal balancing recognizes the separateness of persons, because in that case there is compensation; benefactor and beneficiary are the same person. By contrast, interpersonal balancing ignores the separateness of persons; benefactor and beneficiary are distinct persons, rendering compensation problematic. Critics of utilitarianism think that the separateness of persons shows that utilitarianism has an inadequate conception of distributive justice—utilitarianism is concerned with producing the best total outcome and is otherwise indifferent to the way in which benefits and burdens are distributed.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

In order to respect the separateness of persons, critics claim, concern for persons must take a *distributed* form in which outcomes can be justified to each individually. One such distributed conception is *contractualism*, which claims that distributions of benefits and harms must be acceptable, in the relevant sense, to *each* affected party. One version of contractualism claims that actions and distributions of benefits and harms are right insofar as they conform to principles that no one can reasonably reject. By giving each person a veto, the contractualist seeks a kind of unanimity, in contrast to (p. 379) the aggregative character of utilitarianism. The interpersonally best option may usually be acceptable to many, but it can fail to be acceptable to each. Contractualists often claim that distributions that are acceptable to each will tend toward egalitarianism.

In this way, the separateness of persons purports to play two different roles in discussions of distributive justice. Its *negative* role is to undermine utilitarianism. Indeed, many critics of utilitarianism suppose that the separateness of persons undermines utilitarianism in virtue of its consequentialist or aggregative structure. If so, the negative import of the separateness of persons is to undermine consequentialism more generally. By contrast, the *positive* role of the separateness of persons is to motivate and support a nonaggregative form of egalitarianism about distributive justice.

This essay reconstructs and assesses the negative and positive significance of the separateness of persons. In section 1, I reconstruct utilitarian and consequentialist essentials, focusing on agent-neutral forms of consequentialism, such as utilitarianism. In section 2, I explain the separateness of persons objection to utilitarianism, its reliance on a *compensation requirement*, and how it is supposed to motivate a more egalitarian form of contractualism. In section 3, I turn to assessing this critique, distinguishing between *nonmoralized* and *moralized* versions of the compensation requirement. Utilitarianism does violate a nonmoralized compensation requirement, but that requirement is too strong. Without additional argument, it's not clear what moral theories and theories of distributive justice satisfy the moralized compensation requirement. In section 4, I examine the contractualist claim that the right way to model the separateness of persons involves *pairwise comparison* of options in a way that is fundamentally *anti-aggregative*, arguing that pairwise comparison is problematic precisely because it prevents us from aggregating comparably urgent moral claims. We need to distinguish between *local* and *distal* aggregation; even if we reject aggregating distal claims, we should permit some kinds of local aggregation. This sort of *selective aggregation* is attractive and stands in contrast to the sort of *unrestricted aggregation* that leads to *repugnant conclusions* in both intrapersonal and interpersonal contexts (section 5). I conclude by reconstructing and assessing two arguments for unrestricted aggregation—a sorites-style argument that appeals to the *irrelevance of small differences* (section 6) and another argument that appeals to the *transitivity* of the <better than> relation (section 7). These arguments defend unrestricted aggregation in both intrapersonal and interpersonal cases. They can be resisted by appeal to discontinuities in the value of lives and the urgency of moral claims, which allows the consequentialist to defend aggregation that is selective, rather than unrestricted.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

1. Utilitarianism and Consequentialism

There are a variety of utilitarian and consequentialist moral theories. For instance, classical hedonistic utilitarianism conceives of the good in terms of pleasure and identifies (p. 380) an agent's duty with maximizing pleasure.[1] This makes the good explanatorily prior to the right insofar as it defines right action in terms of promoting the good (cf. Rawls 1971, §5). Generalizing, we might understand consequentialism as the set of moral theories that make the good explanatorily primary, explaining other moral notions, such as duty or virtue, in terms of promoting value. For instance, a consequentialist conception of duty might identify an agent's duty as an action that promotes the good, whereas a consequentialist conception of virtue might identify virtuous dispositions as those with good consequences.

If consequentialism takes the good to be primary and identifies right action as action that promotes value, it contrasts with two different conceptions of right action. *Deontology* takes right action to be the primary normative notion; it recognizes various actions as obligatory, prohibited, or permitted based on their natures and independently of the value they produce. *Virtue ethics* takes the idea of a morally good character to be explanatorily primary in the account of right action; right action, on this view, is action performed by someone with a virtuous character or that expresses a virtuous character (e.g., Watson 1990).

On this way of thinking, consequentialist views treat the good as prior to the right and direct agents to promote the good. Different consequentialist conceptions make different claims about the good, that is, what is intrinsically or nonderivatively good. One issue is whether all goods are *personal* (good for someone) or whether some or all goods are *impersonal* (good independently of any contribution they make to valuable lives). Utilitarian theories are consequentialist theories that take the good to be promoted to be human (or sentient) happiness or well-being.[2] Utilitarian conceptions assume that all goods are personal. Different conceptions of utilitarianism make different claims about the nature of happiness or well-being. One familiar conception is the *hedonistic* claim that pleasure is the one and only intrinsic good and that pain is the one and only intrinsic evil. Alternatively, one might understand the human good in *preference-satisfaction* terms, as consisting in the satisfaction of actual or suitably informed or idealized desire. Hedonism and preference-satisfaction views construe the human good as consisting in or depending upon an individual's contingent and variable psychological (p. 381) states. By contrast, one might understand the good in more objective terms, either as consisting in the *perfection* of one's essential capacities (e.g., one's rational capacities) or as consisting in some *list* of disparate objective goods (e.g., knowledge, beauty, achievement, friendship, or equality).

Another issue the consequentialist must address is *whose* good matters and *how*. Who should an agent care about, and among those that demand her concern, should they matter equally? At one extreme lies the *impartial* consequentialist view that an agent should be concerned to promote any and all kinds of value and, in particular, should have an equal concern with the well-being of all and should act in ways that benefit all those that

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

it is in her power to affect for better or worse. Utilitarianism is probably the most familiar form of impartial consequentialism. It instructs agents to promote human or sentient happiness generally. But a view that recognized impersonal values and instructed agents to promote these wherever possible would also be a form of impartial consequentialism. At the other extreme lies the *partial* consequentialist view that an agent should be intrinsically concerned with promoting only her own happiness or welfare. Such a view would be a form of ethical egoism. In between these extremes lie *moderate* forms of consequentialism that demand nonderivative concern for others but that limit the scope or weight of such concern. One such moderate view is the view that C. D. Broad called "self-referential altruism" and associated with common-sense morality (1953, esp. 279). Self-referential altruism claims that an agent's concerns should have wide scope, but variable weight. It says that an agent has an obligation to be concerned about anyone but that the weight of an agent's moral reasons is a function of the nature of the relationship in which the agent stands to potential beneficiaries. On this view, an agent has reason to be concerned for perfect strangers as well as intimate associates, but, all else being equal, she has more reason to be concerned about the well-being of an associate than a stranger.

These distinctions within consequentialism can also be made in terms of the distinction between agent-neutral and agent-relative reasons. The general form of *agent-relative* reasons makes essential reference to the agent in some way, whereas the general form of *agent-neutral* reasons does not (cf. Nagel 1986, 152). Being under a duty to help children, as such, would involve an agent-neutral reason, whereas being under a duty to help one's own children would involve an agent-relative reason. Being under a duty to minimize suffering would be an agent-neutral reason, whereas the deontological duty never to be the cause of another's suffering, even if this is necessary to minimize total suffering, would be an agent-relative reason. These contrasts can also be captured in the contrast between *constraints* and *options* (e.g., Kagan 1998, chap. 1). Constraints are *categorical moral prohibitions* that are often thought to correlate with moral entitlements that individuals possess—such as *rights*—that limit what someone may do to them, even in the pursuit of good consequences. On such views, it can be wrong to do something, even though doing so might have the best consequences. By contrast, options *permit* an agent to devote attention and resources to her own projects and those of others with whom she is associated out of proportion to their agent-neutral value. Utilitarian and **(p. 382)** agent-neutral conceptions of consequentialism are apparently hostile to both constraints and options, claiming that agents should always do the most good.[3]

Though it is common to associate consequentialism with agent-neutral consequentialism, and utilitarianism is an agent-neutral form of consequentialism, it would be a mistake to treat all consequentialist doctrines as agent-neutral. Ethical egoism and self-referential altruism both identify an agent's duty with promoting goods, though they limit the scope or vary the weight of the values she ought to promote.

Consequentialists have a distinctive orientation or response to values—they promote, rather than honor, the relevant values. To *honor* a value is to represent it as a constraint on action, acting on it or protecting it at every opportunity. To *promote* a value is to take

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

steps that lead to its greater realization overall. But promoting a value overall can require failing to honor it on some occasions, as it would, for example, if promoting and protecting freedom within a community required establishing a compulsory draft. And honoring a value on some occasion may involve failing to promote that value, as it would, for example, if saving an innocent life now could only be done in ways that prevented saving even more innocent lives at some later point in time. Whereas the consequentialist tells agents to promote the relevant values, the deontologist tells them to honor those values (cf. Pettit 1991).

Though more could be said about the forms and limits of consequentialism (see, e.g., Brink 2006), this statement of utilitarian and consequentialist essentials should be adequate for present purposes. On this traditional understanding, consequentialists treat the good as prior to the right and direct agents to promote the relevant goods.[4] Though consequentialism per se is not agent-neutral, many conceptions of consequentialism are agent-neutral. In particular, utilitarianism, whether hedonistic or not, is an agent-neutral form of consequentialism that directs agents to promote (e.g., maximize) human happiness or well-being. Concerns about the separateness of persons target utilitarianism and, more generally, agent-neutral conceptions of consequentialism.

2. The Separateness of Persons

Critics of utilitarianism often focus on its *aggregative* character—the fact that it takes everyone's interests into account, weighs them equally, and balances benefits to some against harms to others, where necessary, so as to produce the best overall outcome. In *A Theory of Justice* (1971), John Rawls famously criticized utilitarianism's aggregative (p. 383) character for ignoring the separateness of persons in the course of defending his own liberal egalitarian conception of justice. In *Anarchy, State, and Utopia* (1974), Robert Nozick likewise criticized utilitarianism's troubles with the separateness of persons in the process of defending his own libertarian entitlement theory of justice. Moreover, both Thomas Nagel (1970) and Bernard Williams (1976) endorsed this criticism of utilitarianism. This consensus about utilitarianism's problems with the separateness of persons is impressive. Both Rawls and Nozick think that these problems reflect utilitarianism's failure to respect individual rights and the demands of distributive justice. Their agreement is all the more striking because they hold such divergent conceptions of rights and justice.

To understand this concern with the separateness of persons, it will help to consider a familiar analogy between prudential aggregation and utilitarian aggregation. Prudence is *temporally neutral* and assigns no intrinsic significance to *when* a benefit or burden occurs within a person's life. It says that we should balance benefits and harms, where necessary, among different stages in a person's life and pursue the action or policy that promotes her overall good best. Utilitarianism is *interpersonally neutral*; it assigns no intrinsic significance to *whom* a benefit or burden befalls. Just as temporal neutrality requires intrapersonal balancing, so too person neutrality requires interpersonal balancing. It re-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

quires that benefits to some be balanced against harms to others, if necessary, to produce the best interpersonal outcome overall. Utilitarianism's person neutrality thus effects a kind of interpersonal balancing akin to the intrapersonal balancing that prudence's temporal neutrality requires.

However, Rawls claims that this sort of interpersonal balancing is unacceptable because it ignores the *separateness of persons*.

> This view of social cooperation [utilitarianism's] is the consequence of extending to society the principle of choice for one man [i.e. prudence], and then, to make this extension work, conflating all persons into one. … Utilitarianism does not take seriously the distinction between persons. (1971, 27)

Nozick, Nagel, and Williams agree. They accept prudence's intrapersonal balancing but reject utilitarianism's interpersonal balancing.

Why accept intrapersonal balancing, but not interpersonal balancing? In *The Methods of Ethics* (1907) Henry Sidgwick suggested we can explain the asymmetry between intrapersonal balancing and interpersonal balancing by appeal to the separateness of persons and the significance of compensation (1907, 418–419, 498). Intrapersonal balancing is compensated, but interpersonal balancing is not.[5] Nozick's discussion develops Sidgwick's thought.

(p. 384)

> Individually, we each sometimes choose to undergo some pain or sacrifice for a greater benefit or to avoid a greater harm. … Why not, *similarly*, hold that some persons have to bear some costs that benefit other persons more? But there is no *social entity* with a good that undergoes some sacrifice for its own good. … To use a person in this way does not sufficiently respect and take account of the fact that he is a separate person, that his is the only life he has. *He* does not get some outweighing good from his sacrifice, and no one is entitled to force this upon him. (1974, 32–33)

Like Sidgwick and others, Nozick is invoking claims about compensation to explain the asymmetric treatment of intrapersonal and interpersonal balancing. Whereas balancing benefits and harms is acceptable *within* a life, balancing benefits and harms *across* lives appears unacceptable. In the intrapersonal case, benefactor and beneficiary are the same person, so compensation is automatic. In the interpersonal case, benefactor and beneficiary are different people; unless the beneficiary reciprocates in some way, the benefactor's sacrifice will not be compensated. Whereas intrapersonal compensation is automatic, interpersonal compensation is not. This fact about compensation appears to rationalize intrapersonal neutrality without rationalizing interpersonal neutrality.

Rawls thinks that utilitarianism's problems with the separateness of persons is symptomatic of its indifference to considerations of individual rights.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> [W]e distinguish as a matter of principle between the claims of liberty and right on the one hand and the desirability of increasing aggregate social welfare on the other; … we give a certain priority, if not absolute weight, to the former. Each member of society is thought to have an inviolability founded on justice or, as some say, natural right, which even the welfare of everyone else cannot override. Justice denies that the loss of freedom for some is made right by the greater good shared by others. The reasoning which balances the gains and losses of different persons as if they were one person is excluded. Therefore in a just society the basic liberties are taken for granted and the rights secured by justice are not subject to political bargaining or to the calculus of social interests. (1971, 27–28)

Similarly, Nozick contrasts goal-based moral theories with constraint-based theories and insists on understanding rights as side-constraints on pursuit of good consequences, rather than as especially important goals (1974, 28–33).

For both Rawls and Nozick, it is impermissible to aggregate the interests of different individuals. Rather, we should respect the separateness of persons by insisting that distributions be acceptable *to each*. Nozick thinks that a distributed concern for each supports an historical entitlement theory of justice in holdings that respects individual talents and choices. A more common reaction is to think that a distributed concern for each supports some form of *contractualism* and its guiding idea that just distributions would be acceptable and justifiable to each and every affected person via the right sort of agreement.

(p. 385) Famously, Rawls appeals to an ex ante contract in the original position, which imposes a veil of ignorance, designed to represent contractors as free and equal moral persons. He claims that contractors in the original position would choose Justice as Fairness, which consists of two principles of justice—a principle of equal basic liberties (Equal Basic Liberties) and a principle that distributes social and economic goods and opportunities so as to be to the greatest benefit of the least advantaged (the Difference Principle). The Difference Principle assigns priority to the worst-off members of society, not permitting greater benefits to the better-off if these do not maximize the prospects of the worst-off. In this way, the Difference Principle is anti-aggregative.

In *The Possibility of Altruism*, Nagel not only criticizes utilitarianism by appeal to the separateness of persons and the compensation principle but also defends Rawls's Difference Principle (1970, 142). Later, in his essay "Equality" Nagel contrasts utilitarianism and egalitarianism. Egalitarianism, he claims, "establishes an order of priority among needs and gives preference to the most urgent, regardless of numbers" (1977, 116–117). Nagel identifies the anti-aggregative Difference Principle as the correct way to give expression to a distributed concern with each affected party.

> So let me return to the issue of unanimity in the assessment of outcomes. The essence of such a criterion is to try in moral assessment to include each person's point of view separately, so as to achieve a result which is in a significant sense acceptable to each person involved or affected. Where there is a conflict of interests, no result can be completely acceptable to everyone. But it is possible to assess

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> each result from each point of view to find the result that is the least unacceptable to the person to whom it is most unacceptable. This means that any other alternative will be more unacceptable to someone than this alternative is to anyone. The preferred alternative is in that sense least unacceptable, considered from each person's point of view separately. A radically egalitarian policy of giving absolute priority to the worst-off, regardless of numbers, would result from always choosing the least unacceptable alternative, in this sense. (1977, 123)

Similarly, in his writings on contractualism, Tim Scanlon thinks that the correct way to model the separateness of persons is via an ex post contract that aims to identify principles that no one can reasonably reject (1982; 1998).

> [Contractualism] holds that an act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behavior that no one could reasonably reject as a basis for informed, unforced general agreement. (1998, 153)

Complaints are the basis of reasonable rejectability, and Scanlon thinks that contractualism forces us to assess complaints of people *individually* (1998, 229–230). This, he thinks, commits us to *pairwise comparison* of the interests and complaints of individuals. So understood, contractualism will register when one individual has a bigger complaint than another, but it will block the aggregation of smaller complaints of several (p. 386) individuals to justify an outcome other than the one that addresses the individually greatest complaints. In this way, Scanlon thinks that contractualism is the correct way to model the separateness of persons and that it contains a restriction to pairwise comparison that makes it fundamentally anti-aggregative.

3. Assessing the Critique of Utilitarianism and Consequentialism

Should we accept the negative thesis associated with the separateness of persons objection? Do utilitarian and consequentialist theories ignore the separateness of persons, and does this show that they lack a plausible conception of distributive justice? There are several issues to explore in addressing these questions.

First, even if utilitarianism countenances interpersonal aggregation and balancing that involves uncompensated sacrifices and so ignores the separateness of persons, this may not be true of all consequentialist theories. In particular, consequentialist theories can assign intrinsic or nonderivative significance to matters of distributive justice, however that is best conceived. Classical utilitarianism demands that we produce the most total happiness, but it seems indifferent how that total is distributed. In particular, it is indifferent between egalitarian and highly inegalitarian distributions of benefits and harms within a population that sum to the same total. Of course, the principle of diminishing marginal utility—the principle that as a person increases her consumption of a given resource, the

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

marginal utility of additional units of that resource decreases—will imply that more equal distributions of resources tend to produce larger amounts of utility. But classical utilitarianism is indifferent between different distributions of the same amount of utility. This is why it is sometimes said that classical utilitarianism has no principle of distributive justice. But if distributive justice is an important good, then consequentialism can insist that it be promoted. Moreover, if the right conception of distributive justice forbids interpersonal aggregation and balancing and the consequentialist assigns distributive justice priority over other goods, then such a consequentialist conception would not be guilty of ignoring the separateness of persons in its conception of distributive justice.

Of course, a consequentialist conception that is sensitive to distributive justice would need to articulate and defend a particular conception of justice as part of its conception of the good. One might wonder if this can be done while maintaining the priority of the good over the right (Rawls 1971, 25). This is a reasonable question. It might seem circular if we define the right in terms of the good and then define the good in terms of permissible distribution. But it's not clear that we couldn't have a theory of the good that is distribution-sensitive and still maintain that the good is prior to the right. If the right is all-things-considered obligation and permission, then it seems we could define the right in terms of promoting the good while allowing that the distribution of goods in a situation is an ingredient, perhaps the most important ingredient, in the overall goodness of that situation.

(p. 387) Second, we might wonder if the separateness of persons objection proves too much. Recall that the objection accepts intrapersonal aggregation and balancing but rejects interpersonal aggregation and balancing, because whereas there is automatic intrapersonal compensation for sacrifice, interpersonal compensation is not automatic. In this way, the objection seems to assume that compensation is both necessary and sufficient for demands for sacrifice to be legitimate. The sufficiency claim is what rationalizes the intrapersonal balancing that prudence requires, and the necessity claim is what undermines the sort of interpersonal balancing that utilitarianism requires.

1. Sacrifice is legitimate just in case it is compensated.
2. Hence, intrapersonal temporal neutrality is required by the sufficiency of compensation for sacrifice; because beneficiary and benefactor are the same in the intrapersonal case, sacrifice is automatically compensated.
3. Hence, the necessity of compensation for sacrifice blocks interpersonal neutrality; because benefactor and beneficiary are distinct in the interpersonal case, sacrifices are not automatically compensated.

But the assumption about legitimate sacrifice in (1) is potentially problematic.

The sufficiency claim is itself interesting. Consider a *now-for-then sacrifice* in which a proximate sacrifice makes possible a greater distal benefit for the agent. We might think that the compensation involved in now-for-then sacrifice gives the agent a prudential reason, perhaps a sufficient one, to make the sacrifice. Nonetheless, concerns about paternalism might make us reluctant to demand of others that they make now-for-then sacri-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

fices. Interesting as the sufficiency thesis is, it's the necessity claim that blocks interpersonal aggregation. But the necessity claim is quite extreme.

One measure of the extremity of the necessity claim is that an *egoist* form of consequentialism could embrace it. For egoism is the view that an agent has reason to do something just insofar as that would advance his own happiness or well-being. As such, egoism embraces both the sufficiency and the necessity of compensation for sacrifice. Egoism is a potentially interesting doctrine, especially as a claim about rationality, rather than morality, worth exploration in another context (e.g., Brink 1997a; 1997b). But ethical egoism seems implausible to many as an account of our duties to others and, hence, as an account of distributive justice. For the necessity of compensation for sacrifice would seem to preclude any uncompensated duties to others. In particular, it implies that it would be impermissible to ask one person to make or even risk a very small sacrifice for the sake of enormous benefits to others. For example, duties of *easy rescue* violate the necessity claim.

Another measure of the extremity of the necessity claim is that many conceptions of distributive justice flout it. Consider Rawls's own Difference Principle, which permits only those inequalities in social and economic goods that are to the greatest benefit of the least advantaged (1971, §§12–13, 21, 26). As Rawls himself recognizes, the Difference Principle will in many circumstances require the better-off to forego further benefits for the sake of the worse-off (1971, 103). These would often be uncompensated sacrifices. But then the Difference Principle also violates the necessity claim and so would offend against the separateness of persons.

(p. 388) The necessity of compensation for sacrifice is overly restrictive if our conception of a sacrifice counts any cost or negative effect on someone's well-being as a sacrifice. For then even easy rescue's demand of small sacrifices from one to avert significant harm or produce significant benefit for others violates necessity. However, we might *moralize sacrifice*, claiming that what's impermissible is not imposing any sacrifice on some for the benefit of others but only *unjustified* sacrifices. This sort of moralized compensation principle is less restrictive and permits some kinds of interpersonal balancing. For instance, it need not condemn requiring easy rescue. Moreover, it need not condemn the sort of interpersonal balancing that the Difference Principle requires. Whereas Rawls condemns the sort of bottom-up sacrifice of the worse-off for the sake of the better-off that he thinks utilitarianism might demand, he defends the sort of top-down sacrifice of the better-off for the sake of the worse-off that the Difference Principle might demand (cf. 1971, 103, 178). In doing so, Rawls seems to accept a *moral asymmetry* between top-down and bottom-up sacrifice—all else being equal, bottom-up sacrifice is morally more problematic than top-down sacrifice. If we combine a moralized compensation principle with the moral asymmetry thesis, we might be able to reconcile the Difference Principle, but not some forms of utilitarianism, with a suitably moralized compensation principle.

But, of course, it is a *substantive* question which sacrifices are justified and which are unjustified and whether some version of the moral asymmetry thesis is true. Once we drop

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

the objection to interpersonal balancing per se, it is no longer clear that all forms of consequentialist balancing are impermissible or that egalitarian distributions are permissible. The unmoralized separateness of persons objection, resulting from an unmoralized compensation principle, proves too much. The moralized separateness of persons objection, resulting from the moralized compensation principle, *poses*, rather than resolves, questions about the role of aggregation within distributive justice.

Rawls, Nozick, and Nagel write as if the bare fact that utilitarianism requires interpersonal balancing and permits uncompensated sacrifices is sufficient reason to conclude that it ignores the separateness of persons and should be rejected, independently of the kind of sacrifices required. This strongly suggests that they originally understand the compensation requirement in a nonmoralized way, which, at least for Rawls and Nagel, is hard to reconcile with their own claims about distributive justice. Reconciliation is possible if they adopt the moralized version of the compensation requirement. But then the case against utilitarianism cannot be made in advance of defending a particular moralization of the compensation requirement and some version of the moral asymmetry thesis. The separateness of persons is a conversation *starter*, not the conversation *stopper*, it is often presented as.

4. Assessing Pairwise Comparison and Anti-Aggregation

Critics of utilitarianism appeal to the separateness of persons not only to criticize utilitarianism but also to motivate alternative distributional principles. As we have seen, (p. 389) Rawls, Nagel, and Scanlon all think that the correct way to model the separateness of persons is contractualist and that the right kind of contract will direct our attention to the worst-off, giving them priority that blocks interpersonal aggregation. Rawls imagines that contractors in the original position compare representative social positions, effectively ignoring the numbers of people occupying such positions, arguably employing pairwise comparisons of representative social positions and preventing the sort of interpersonal aggregation that utilitarianism permits. Both Nagel and Scanlon explicitly claim that in assessing outcomes under different principles we should restrict ourselves to pairwise comparison, effectively blocking interpersonal aggregation.

In assessing these proposals, we should distinguish two different commitments that these contractualist egalitarians make—*prioritarianism* and *anti-aggregation*. Prioritarianism reflects the operation of moral asymmetry—the thesis that, all else being equal, the worse-off have more urgent claims than the better-off. Anti-aggregation is embodied in the restriction to individual or pairwise comparison, preventing a larger number of smaller individual claims from outweighing a smaller number of individually larger claims. These two theses need not go together. In particular, it is possible to embrace moral asymmetry and prioritarianism without embracing anti-aggregation (see, e.g., Brink

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

1993; 2015; and Arneson 2000). While some forms of aggregation appear morally problematic, the wholesale rejection of aggregation is also problematic.

Let's agree, at least for the sake of argument, that the moral asymmetry thesis is correct and that, as a result, some form of prioritarianism is also correct. We need to specify the dimensions along which we determine how well or badly off someone is and how serious their complaints about some options are. I will assume that how well off someone is a function of both her absolute and relative well-being. Though A is worse off than B in virtue of A's relative level of well-being, we are only likely to think this is morally important if A's absolute level of well-being is low. If both A and B are fabulously well off, we are unlikely to care much about A's purely relative deprivation. Contractualist discussions often focus on the complaints of affected parties. Scanlonian contractualism is explicit about this. We can measure the size of someone's complaint about one option (e.g., principle or outcome) by how much better she would have fared under some alternative option and by her relative level of well-being. No doubt, there is more to be said about how to measure how well-off people are and the size of their complaints. However, these remarks should be sufficient for present purposes.[6]

Prioritarianism says that all else being equal, we should prefer options that favor the worse-off. Rawls's Difference Principle, which requires foregoing benefits to the better-off so that the position of the worst-off can be maximized, is a form of prioritarianism. Nagel illustrates prioritarianism with an example in which someone has two children, one with special needs. Treating them with equal concern, Nagel claims, requires unequal (p. 390) treatment; specifically, it requires devoting greater attention and resources to the child with special needs (1977, 124). When other things are equal, moral asymmetry supports prioritarian conclusions in which we should benefit the worst-off.

But other things are not equal if we vary the numbers of affected parties. Consider a case in which there are two people, A and B, who are both badly off, but A is marginally worse-off than B. If we had one indivisible resource to distribute, prioritarianism would favor giving it to A. So far, so good. But now assume that we have two groups of similarly situated people. Both the As and the Bs are badly off, but the As are marginally worse off. But now assume that there are many more Bs than As. Do the individually slightly more urgent claims of As take priority over individually slightly less urgent but vastly more numerous claims of the Bs? Pairwise comparison and anti-aggregation assume so. But in many cases that would be an implausible answer.

Consider two examples. The first is a choice of educational policy. Assume that we have a fixed number of resources to devote to special education. Assume that we can quantify the severity of learning disabilities and the amount of benefit that different policies would confer. One disability (A) is marginally more severe but also quite rare. Because this disability is more severe, it is harder to overcome than the other (B), and so education of these A-children is more resource-intensive. One policy (Hardship) gives priority to those with the individually worse disability, while the other (Benefit) gives priority to the much larger group with the marginally less severe disability. The second case involves analo-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

gous issues about how to ration scarce healthcare resources. Assume that we have two conditions, A and B. Those with condition A are marginally worse-off individually than those with condition B, but whereas condition A is quite rare, condition B is quite common. Condition A is harder to treat, so we have to choose between Hardship, which provides a small benefit in terms of life expectancy to A-patients, and Benefit, which produces a larger benefit in terms of life expectancy to the much larger group of B-patients.

In both cases, pairwise comparison requires that we prefer Hardship to Benefit, no matter how few A-type individuals there are and how many B-type individuals there are. That strikes me as unreasonable. We can embrace moral asymmetry and prioritarianism without endorsing pairwise comparison and its ban on all forms of interpersonal aggregation.

5. Unrestricted Aggregation?

Scanlon recognizes this sort of worry about his requirement of pairwise and its wholesale ban on interpersonal aggregation, but he doesn't see a way to allow interpersonal aggregation only selectively and in the right way (1998, 230–241). Intuitively, not all aggregation is morally equal. Egalitarian and prioritarian critics of classical utilitarianism rightly worry about unrestricted interpersonal aggregation that would impose serious harms on a few for the sake of preventing significantly smaller harms or producing modest benefits to a much larger number of persons. The worry is strongest where the goods and harms being aggregated are—considered individually—of very *disparate* value. We should be reluctant to think that even one person should lose her life even if (p. 391) this meant that we could prevent millions or billions of hangnails. More realistically, we should be reluctant to think that small numbers should be imprisoned or denied basic liberties, so that a much larger number of people could avoid offense or inconvenience. These intuitions are amplified when the few are already worse-off than the many. It would be good to block that kind of interpersonal aggregation. But a wholesale ban on interpersonal aggregation also blocks legitimate forms of interpersonal aggregation, where the kind and magnitude of benefits and harms are—considered individually—*similar*, rather than disparate. In such cases, the ban on interpersonal aggregation looks problematic. To explain why we should prefer Benefit to Hardship, we need to allow interpersonal balancing across individually similar cases.

What we need is a way of blocking aggregation where *the stakes are highly disparate* and permitting it where *the stakes are roughly similar*. This requires an account of when the stakes are disparate or similar, and there are different possible views about the relevant dimensions of similarity and disparity. For present purposes, let's assume that similarity/disparity is a function of three different dimensions: (a) the *kinds* of goods and harms in question, (b) their *magnitude*, and (c) the *relative position* of the affected parties. Presumably, it's clear how the kind and magnitude of goods and bads can affect whether the stakes are similar or not. Moral asymmetry would also imply that whether the stakes are morally similar depends on the relative position of the affected parties. The best-off and the worst-off in a highly inegalitarian society occupy disparate social positions, whereas

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

adjacent social positions in a social hierarchy will have similar standing. Importantly, each of these dimensions is *scalar*, forming a continuous scale and ensuring that there will be cases in which it is *indeterminate* whether the stakes are similar or not.

One response to the discussion so far would be to defend *selective*, rather than *unrestricted* aggregation. In particular, one might seek to defend aggregation in cases involving similar stakes, but reject it in cases involving highly disparate stakes. On such a view, we would defend *local aggregation* but reject *distal aggregation*. However, there is an obstacle to defending this sort of selective aggregation. There are arguments for unrestricted aggregation. We can distinguish the operation of unrestricted aggregation in interpersonal and intrapersonal cases.

Derek Parfit's *repugnant conclusion* is a consequence of unrestricted interpersonal aggregation (1984, chap. 17). The repugnant conclusion claims that there is a world with an extremely large population of people with lives barely worth living that is better than a world with a large population of people leading exceptionally good lives. Parfit reaches the repugnant conclusion by a long sequence of small trade-offs of a larger number of slightly lesser goods for a smaller number of slightly superior goods. Similarly, this reasoning supports forms of unrestricted interpersonal aggregation that imply that we should accept trade-offs in which we prevent some very large number of hangnails at the cost of someone's life or in which we sacrifice one person's basic liberties for the sake of preventing offense to a great many people.[7]

(p. 392) We can also imagine unrestricted *intrapersonal* aggregation. In the *Philebus*, Plato contrasts the life of intelligence and a life with only the simple pleasures of an oyster, defending the superiority of a life of intelligence against an oyster-like existence (20b5–22e5). In *The Nature of Existence*, J. M. E. McTaggart argues that even if the goods of the intellect are much superior to the value of the oyster's humble pleasures—such that, all else being equal, we should strongly prefer the life of the intellect to the life of the oyster—nonetheless we have reason to prefer the life of an oyster provided that is sufficiently longer than the life of the intellect (1927, vol. II, book vii, chap. 7, §§868–870).[8] But many would disagree, regarding this, as McTaggart notes, as a "repugnant conclusion." Presumably, Plato would. In *Utilitarianism*, John Stuart Mill defends the higher pleasures doctrine, claiming that a life containing the higher pleasures (Socrates dissatisfied) is *discontinuously better* than a life containing only lower pleasures (the pig or fool satisfied) (1861, chap. II, paras. 3–6).[9] But one might defend McTaggart's conclusion indirectly, rather than directly, by appeal to a sequence of local intrapersonal comparisons.

There are at least two versions of this kind of argument for unrestricted aggregation worth distinguishing. Both versions begin from the assumption of selective local aggregation—the idea that a significantly larger number of individually slightly less important claims can outweigh a smaller number of claims that are individually slightly more important—and progress by a sequence of cases to distal aggregation—the idea that an enormous number of extremely unimportant claims can outweigh a small number of extremely important claims. In the first version, the disparity among lives or goods compared is

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

gradually widened, moving from local comparisons to progressively more distal comparisons, and there seems no nonarbitrary point at which to say that we shouldn't prefer a greater number of lesser goods or claims. In the second version, we only make local comparisons. But the transitivity of the <better than> relation uses a series of local comparisons to commit us to distal aggregation.

6. A Sorites Argument for Unrestricted Aggregation

Here's one attempt to formulate more carefully the first version of the argument for unrestricted aggregation, whether intrapersonal or interpersonal.

1. *Sequence*. Consider a sequence of lives or worlds A-Z. The goodness of lives in these worlds is a function of the kinds and magnitudes of goods people enjoy and whether they have any complaints and how serious they are. In the intrapersonal sequence: A is a full life filled with all the most important goods; B is (p. 393) almost as good as A, but significantly longer; C is almost as good as B but significantly longer; and so on until Z, which is barely worth living but fantastically long. In the interpersonal sequence: A is a world in which there are many people leading very good lives; B is a world in which everyone is leading lives almost as good as in A, but there are many more people; C is a world in which everyone is leading lives almost as good as in B, but there are many more people; and so on until Z, which is a world with an enormous population all leading lives barely worth living.
2. *Local Aggregation*. Aggregation is permissible in cases involving similar goods and bads. In particular, all else being equal, B is better than A; C is better than B; and so on, including the fact that Z is better than Y.
3. *Irrelevance of Small Differences*. Small differences can't make for big moral or evaluative differences. In particular, if B is better than A, then, given the small differences between B and C, C must also be better than A; if D is better than C, then, given the small differences between C and D, D must also be better than A; and so on.
4. Hence, *Unrestricted Aggregation*. Z is better than A—The intrapersonal version: A sufficiently long life that is never more than barely worth living is better than a long life filled with the greatest goods; the interpersonal version: a world with an enormous population of people all leading lives barely worth living is better than a world in which there are many people leading very good lives.

This defense of unrestricted aggregation seems relevantly like a sorites argument inasmuch as it relies on the irrelevance of small differences claim. For this reason, we might call it the sorites version of unrestricted aggregation. The original sorites argument is a step-wise argument designed to convince us that there are no such things as heaps of sand. One grain of sand is not a heap. Adding one grain of sand to something that is not a heap cannot produce something that is a heap. As we add grains of sand, there is no nonarbitrary point at which a heap emerges. But then it follows that there is no number of grains of sand, however large, that constitutes a heap of sand. This argument for unre-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

stricted aggregation may seem to have a similar structure. It takes us from the assumption of local aggregation to a conclusion about distal aggregation by gradually moving from local comparisons to progressively more distal comparisons and insisting that there is no nonarbitrary point at which to say that we shouldn't prefer a sufficiently greater number of lesser goods or lives.

Should we accept this argument for unrestricted aggregation? Its similarity to sorites arguments might make us suspicious. Though the exact diagnosis of where the sorites argument goes wrong is controversial, it is widely viewed as fallacious, because it begins from a true premise and leads to a manifestly false conclusion. One sign that something is fishy about the sorites argument is that it is *reversible*. One version of sorites begins by assuming that one grain of sand is not a heap and then appeals to successive applications of the claim that small differences are irrelevant to show that there is no number of grains of sand that would constitute a heap. But the same logic proves that any number of grains of sand, no matter how few, is a heap provided we begin by assuming that very large numbers of grains of sand constitute a heap. This version of sorites begins by (p. 394) assuming that a great many grains of sand is a heap and then appeals to successive applications of the claim that small differences are irrelevant to show that any number of grains of sand, no matter how few, is a heap. Similarly, the sorites version of the unrestricted aggregation argument is reversible and will support a complete ban on aggregation if only we begin with the assumption that distal aggregation is impermissible. For then by appeal to successive applications of the irrelevance of small differences, we reach the conclusion that even local aggregation is impermissible.

We began with the assumption of selective aggregation, in particular, that while local aggregation is permissible distal aggregation is not. But then the reversibility of the sorites argument for unrestricted aggregation is problematic, because the two starting points of the argument lead to incompatible conclusions. This suggests that the sorites argument is problematic. But how? Sequence is just a statement of the structure of the situation under investigation, and local aggregation is an assumption common to both selective and unrestricted aggregation. So the only remaining premise is the claim about the irrelevance of small differences. This principle is not problematic in itself. What's problematic is its *repeated* application within the sequence. For repeated small differences amount to a *large difference*, and large differences are relevant.

What would it mean to reject the repeated application of the principle about the irrelevance of small differences? One might think that the impermissibility of aggregation grows as the small differences accumulate. The first local aggregation involved in B being better than A is permissible. Indeed, the first few small differences may be irrelevant to the permissibility of aggregation. But as the differences accumulate, the case against aggregation grows until it becomes impermissible. On one version of this view, there is some nonarbitrary point in the sequence at which a small additional difference renders what was previously acceptable aggregation suddenly unacceptable. Perhaps B-L is better than A, but M-Z is not. But it's hard to see what would justify singling out any one small difference to make this big difference. A more plausible version of this view, I think,

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

claims that there is determinate permissibility at the beginning of the sequence (e.g., B is determinately better than A), determinate impermissibility at the end of the sequence (e.g., Z is determinately not better than A), and indeterminate permissibility in some large range of cases in the middle. Of course, if this alternative picture is to avoid positing its own arbitrary point at which small differences make the difference between permissibility and indeterminate permissibility, on the one hand, and between indeterminate permissibility and determinate impermissibility, on the other hand, there will need to be higher-order indeterminacy about when first-order indeterminacy sets in and then stops. I think that this is an attractive view about how to respond to many sorites-style arguments and how to preserve selective aggregation in the face of this sorites argument for unrestricted aggregation.[10]

(p. 395) 7. A Transitivity Argument for Unrestricted Aggregation

In the sorites argument for unrestricted aggregation, the disparity among lives or goods compared is gradually widened, moving from local comparisons to progressively more distal comparisons, and there seems no nonarbitrary point at which to say that we shouldn't prefer a greater number of lesser goods or lives. By contrast, in the transitivity argument for unrestricted aggregation, we only make local comparisons, but the transitivity of the <better than> relation uses a series of local comparisons to commit us to distal comparisons.

The transitivity argument begins from the assumption of selective aggregation—the idea that a significantly larger number of individually slightly lesser goods or claims can morally outweigh a smaller number of goods or claims that are individually slightly more important. But the process of local aggregation can repeat indefinitely. As long as the <better than> relation is transitive, this implies unrestricted aggregation—the idea that a sufficiently large number of de minimus goods or claims can be better than a much smaller number of claims that are individually much more important. Stated more carefully, the argument has a structure something like this:

1. *Sequence*. Consider a sequence of lives or worlds A-Z. The goodness of lives in these worlds is a function of the kinds and magnitudes of goods people enjoy and whether they have any complaints and how serious they are. In the intrapersonal sequence: A is a full life filled with all the most important goods; B is almost as good as A, but significantly longer; C is almost as good as B but significantly longer; and so on until Z, which is barely worth living but fantastically long. In the interpersonal sequence: A is a world in which there are many people leading very good lives; B is a world in which everyone is leading lives almost as good as in A, but there are many more people; C is a world in which everyone is leading lives almost as good as in B, but there are many more people; and so on until Z, which is a world with an enormous population all leading lives barely worth living.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

2. *Local Aggregation*. Aggregation is permissible in cases involving similar goods and bads. In particular, all else being equal, B is better than A; C is better than B; and so on, including the fact that Z is better than Y.
3. *Transitivity*. The relation <better than> is *transitive*—If B is better than A, and C is better than B, then C is better than A.
4. Hence, *Unrestricted Aggregation*. Z is better than A—The intrapersonal version: A sufficiently long life that is never more than barely worth living is better than a long life filled with the greatest goods; the interpersonal version: a world with an enormous population of people all leading lives barely worth living is better than a world in which there are many people leading very good lives.

(p. 396) The crucial difference between the sorites and transitivity arguments for unrestricted aggregation is that the former relies on the iteration of the irrelevance of small differences principle, whereas the latter relies on transitivity.

Parfit represents the repugnant conclusion as following from this kind of transitivity argument.[11] The transitivity argument for unrestricted aggregation may seem more robust than the sorites argument, precisely because transitivity may seem more plausible than the irrelevance of small differences. Should we accept the transitivity argument? The transitivity argument seems valid. To avoid unrestricted aggregation, one must reject either local aggregation or transitivity.

Consider local aggregation. Notice that local aggregation only compares adjacent pairs—A and B, B and C, and so on, including Y and Z. Unlike the sorites version, the transitivity version does not assume that we can compare A and C, A and D, and so on. It derives such comparisons from local aggregation and transitivity. Perhaps local aggregation fails at some point between adjacent lives and worlds in the sequence. But since the space between adjacent pairs is by hypothesis minimal, if local aggregation between some pairs is permissible, it's unclear why it wouldn't be permissible in all cases involving adjacent pairs.

Consider transitivity. We might think that rational decision-making depends on the assumption of transitivity. If transitivity never obtained, we could never do more than pairwise comparison, which would severely limit our ability to compare values and construct preference orderings. But there's no need to deny *local transitivity*. The issue is really about whether we need to accept *distal transitivity*. In particular, the issue is whether transitivity holds in cases in comparisons involving significantly different *kinds* of value. In the intrapersonal case, the issue is whether transitivity holds in cases involving larger and smaller numbers of very different kinds of goods, such as higher and lower pleasures or perfection and contentment. In the interpersonal case, the issue is whether transitivity holds in cases involving larger and smaller numbers of people possessing very different kinds of goods and also occupying significantly different positions in a social hierarchy. As Mill argues in the intrapersonal case and as Rawls argues in the interpersonal case, we might want to recognize comparative *discontinuities* in kinds of goods and social position. One kind of discontinuity is *lexical priority* in which no amount of a lesser kind of life is

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

better than the smallest increment in the best kind of life or in which no amount of trivial benefits to the better-off is better than the greatest kind of benefits to the worst-off. If we embrace some kinds of discontinuity, we can deny distal transitivity involving comparisons between highly disparate kinds of goods and lives.

Where exactly along the local/distal spectrum does transitivity cease to obtain? Despite differences between the sorites and transitivity arguments, I think they are similar on this point. It could be that there is some precise point in applying the transitivity (p. 397) principle to ever more distal comparisons that transitivity fails. Maybe there is some small change in the kinds of value within a life or the proximity of classes in a social hierarchy that blocks aggregative comparisons. Perhaps transitivity holds in cases involving comparisons A-L, but not involving comparisons between A-L and M-Z. But it's hard to see what would justify singling out any one small difference to make this big difference. A more plausible claim, I think, is that transitivity is itself scalar and becomes more problematic as it becomes more distal. This would mean that transitivity determinately obtains in cases of local comparison, determinately fails to obtain in cases of distal comparison, and is indeterminate in its application to a range of comparisons in the middle. Of course, if this alternative picture is to avoid positing its own potentially arbitrary points at which small differences make the difference between transitivity and indeterminate transitivity, on the one hand, and between indeterminate transitivity and determinate intransitivity, on the other hand, there will need to be higher-order indeterminacy about when first-order indeterminacy sets in and then stops.

This way of avoiding unrestricted aggregation and the intrapersonal and interpersonal repugnant conclusions requires denying that transitivity holds everywhere.[12] This will prevent a complete ordering of goods, but a complete ordering may be both unnecessary and misguided (e.g., Sen 1973, 4–6, 47–76).

8. Concluding Remarks

It is time to take stock. Utilitarianism is an agent-neutral form of consequentialism that takes the good to be promoted to be happiness or well-being. The separateness of persons objection alleges that there is an asymmetry between the sort of intrapersonal balancing of benefits and harms within a life that prudence requires and the sort of interpersonal balancing of benefits and harms across lives that utilitarianism and other forms of agent-neutral consequentialism require. This argument rests on the assumption that sacrifice requires compensation, because intrapersonal compensation is automatic in a way that interpersonal compensation is not. Even if we accepted the separateness of persons argument against utilitarianism, it would not be successful against other consequentialist conceptions that took the appropriate form of distributive justice to be an important intrinsic or nonderivative good. Moreover, as it stands, the separateness of persons argument is inconclusive. If we rely on a nonmoralized conception of compensation, then the compensation requirement is much too extreme, undermining any theory that demands even de minimus sacrifices by some for the sake of great benefits to others, as in easy

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

rescue. The compensation require (p. 398) ment is more plausible if we moralize compensation so that only unjustified sacrifices are prohibited and recognize a moral asymmetry between the claims of the better-off and worse-off. But it is a substantive question involving independent argument that utilitarianism and consequentialism cannot satisfy the moralized compensation requirement. Contractualist critics of utilitarianism claim that the right way to model the separateness of persons involves pairwise comparison of options in a way that is fundamentally anti-aggregative. But pairwise comparison is problematic precisely because it prevents us from aggregating comparably urgent moral claims, as in the examples involving Hardship and Benefit. We need to distinguish between local and distal aggregation; even if we reject aggregating distal claims, we should permit some kinds of local aggregation. This sort of selective aggregation is attractive and stands in contrast to the sort of unrestricted aggregation that leads to repugnant conclusions in both intrapersonal and interpersonal contexts. We can distinguish two arguments for unrestricted aggregation, whether intrapersonal or interpersonal—a sorites-style argument that appeals to the irrelevance of small differences and another argument that appeals to the transitivity of the <better than> relation. These arguments can be resisted by appeal to discontinuities in the value of lives and the urgency of complaints, which allows the consequentialist to defend aggregation that is selective, rather than unrestricted.[13]

References

Alexander, Larry. 2000. “Deontology at the Threshold.” *San Diego Law Review* 37: 893–912.

Arneson, Richard. 2000. “Luck Egalitarianism and Prioritarianism.” *Ethics* 110: 339–349.

Barnes, Elizabeth. 2010. “Ontic Vagueness: A Guide for the Perplexed.” *Noûs* 44: 601–627.

Barnes, Elizabeth. 2014. “Fundamental Indeterminacy.” *Analytic Philosophy* 55: 339–362.

Brink, David. 1993. “The Separateness of Persons, Distributive Norms, and Moral Theory.” In *Value, Welfare, and Morality*, edited by C. Morris and R. Frey. New York: Cambridge University Press.

Brink, David. 1997a. “Rational Egoism and the Separateness of Persons.” In *Reading Parfit*, edited by J. Dancy. Oxford: Blackwell.

Brink, David. 1997b. “Self-love and Altruism.” *Social Philosophy & Policy* 14: 122–157.

Brink, David. 2006. “Some Forms and Limits of Consequentialism.” In *The Oxford Handbook in Ethical Theory*, edited by D. Copp. Oxford: Clarendon Press.

Brink, David. 2013. *Mill’s Progressive Principles.* Oxford: Clarendon Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Brink, David. 2015. “Justice as Fairness, Utilitarianism, and Mixed Conceptions.” In *The Original Position*, edited by T. Hinton. Cambridge: Cambridge University Press.

Broad, C. D. 1953. “Self and Others.” Reprinted in *Broad’s Critical Essays in Moral Philosophy*, edited by D. Cheney. London: George Allen & Unwin, 1971.

Carlson, Erik. 2000. “Aggregating Harms—Should We Kill to Avoid Headaches?” *Theoria* 66: 246–255.

(p. 399) Hyde, Dominic. 2018. “Sorites Paradox.” In *Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. **https://plato.stanford.edu/entries/sorites-paradox**.

Kagan, Shelly. 1998. *Normative Ethics.* Boulder, CO: Westview.

Kraut, Richard. 2018. *The Quality of Life.* Oxford: Clarendon Press.

Li, Hon Lam. Unpublished manuscript. “Contractualism and Aggregation.”

McKerlie, Dennis. 1989. “Equality and Time.” *Ethics* 99: 475–491.

McTaggart, J. M. E. 1927. *The Nature of Existence*, 2 vols. Cambridge: Cambridge University Press.

Mill, John Stuart. (1861). 1998. *Utilitarianism*. Edited by R. Crisp. Oxford: Oxford Classical Texts.

Moore, Michael. 1997. *Placing Blame*. Oxford: Clarendon Press.

Nagel, Thomas. 1970. *The Possibility of Altruism.* Oxford: Clarendon Press.

Nagel, Thomas. 1972. “War and Massacre.” *Philosophy & Public Affairs.* Reprinted in Nagel (1979).

Nagel, Thomas. 1977. “Equality.” Reprinted in Nagel (1979).

Nagel, Thomas. 1979. *Mortal Questions.* New York: Cambridge University Press.

Nagel Thomas. 1986. *The View from Nowhere.* New York: Oxford University Press.

Nebel, Jacob. 2018. “The Good, the Bad, and the Transitivity of Better Than.” *Noûs* 52: 874–899.

Norcross, Alastair. 1997. “Comparing Harms: Headaches and Human Lives.” *Philosophy & Public Affairs* 26: 135–167.

Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.

Parfit, Derek. 1984. *Reasons and Persons.* Oxford: Clarendon Press.

Parfit, Derek. 2016. “Can We Avoid the Repugnant Conclusion?” *Theoria* 82: 110–127.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Pettit, Philip. 1991. “Consequentialism.” Reprinted in *Consequentialism*, edited by S. Darwall. Oxford: Blackwell, 2003.

Plato. *Philebus.* Translated by J. Gosling. Oxford: Clarendon Plato Series, 1975.

Portmore, Douglas. 2011. *Commonsense Consequentialism.* Oxford: Clarendon Press.

Pummer, Theron. 2017. “Spectrum Arguments and Hypersensitivity.” *Philosophical Studies* 75: 1729–1744.

Pummer, Theron. Forthcoming. “Sorites on What Matters.” In *Essays in Honour of Derek Parfit, vol. I: Normative Ethics and Personal Identity,* edited by T. Campbell, J. McMahan, and K. Ramakrishnan. Oxford: Clarendon Press.

Rachels, Stuart. 1998. “Counterexamples to the Transitivity of Better Than.” *Australasian Journal of Philosophy* 76: 71–83.

Rachels, Stuart. 2004. “Repugnance or Intransitivity: A Repugnant but Forced Choice.” In *The Repugnant Conclusion: Essays on Population Ethics*, edited by J. Ryberg and T. Tannsjo. Dordrecht: Kluwer.

Rawls, John. 1971. *A Theory of Justice.* Cambridge, MA: Harvard University Press.

Scanlon, T. M. 1982. “Contractualism and Utilitarianism.” In *Utilitarianism and Beyond*, edited by A. Sen and B. Williams. Cambridge: Cambridge University Press.

Scanlon, T. M. 1998. *What We Owe to Each Other.* Cambridge, MA: Harvard University Press.

Sen, Amartya. 1973. *On Economic Inequality.* Oxford: Clarendon Press.

Sidgwick, Henry. 1907. *The Methods of Ethics*. 7th ed. New York: Macmillan.

Sorensen, Roy. 2018. “Vagueness.” In *Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. **https://plato.stanford.edu/entries/vagueness**.

(p. 400) Temkin, Larry. 1996. “A Continuum Argument for Intransitivity.” *Philosophy & Public Affairs* 25: 175–210.

Temkin, Larry. 2012. *Rethinking the Good.* Oxford: Clarendon Press.

Watson, Gary. 1990. “On the Primacy of Character.” In *Identity, Character, and Morality.* Cambridge, MA: MIT Press.

Williams, Bernard. 1976. “Persons, Character, and Morality.” Reprinted in Bernard Williams, *Moral Luck.* Cambridge: Cambridge University Press, 1981.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Notes:

([1]) Different conceptions of pleasure are possible. Some conceive of pleasure in phenomenal terms, as having a common kind of feel or qualia that varies principally in intensity and duration. Others understand pleasure in functional terms, taking it to be a mental state or sensation that the subject likes and is disposed, other things being equal, to prolong. My focus will be on utilitarianism per se, rather than hedonistic utilitarianism, so present purposes do not require adopting a particular conception of happiness or pleasure.

([2]) Many classical hedonistic utilitarians, such as Jeremy Bentham and Henry Sidgwick, conceive of the good as *sentient*, and not just human, happiness or pleasure. Whether to focus on human or sentient happiness or pleasure is an important choice point for utilitarians with significant practical implications. However, it will be simpler, in the issues about distributive justice that I will be discussing, to focus on the special case of distribution among humans, even if we eventually decide we need to fold sentient interests into our conception of distributive justice.

([3]) A sensible deontological morality employs constraints and options that are *moderate,* rather than *absolute*—prohibitions and permissions that are overridden if the cost of observing them is sufficiently great. For discussion, see Nagel (1972); Nozick (1974, 30n); Moore (1997, 721–725); Kagan (1998, 79); and Alexander (2000).

([4]) For a different and potentially more ecumenical conception of consequentialism, see, e.g., Portmore (2011).

([5]) On the one hand, Sidgwick is attracted to the utilitarian extension of balancing goods and harms from the intertemporal prudential case to the interpersonal case. On the other hand, Sidgwick thinks that the separateness of persons and compensation provide a rationale for the egoist to resist the utilitarian conclusion. This ambivalence is reflected in the fact that Sidgwick reluctantly concludes *The Methods of Ethics* by recognizing a dualism of practical reason between agent-relative and agent-neutral methods of ethics—egoism and utilitarianism.

([6]) Like others, I will assume that we assess options and complaints about the options in terms of the effects of the options on the overall life prospects of affected parties, not the effects on temporal parts of their lives. So, for instance, the Difference Principle requires maximizing the life prospects of the worst-off person, not minimizing the worst period in anyone’s life. For exploration of this assumption, see, e.g., McKerlie (1989).

([7]) There is a complex literature discussing unrestricted interpersonal aggregation, which includes Parfit (1984, chap. 17; 2016); Temkin (1996; 2012); Norcross (1997); Rachels (1998; 2004); Carlson (2000); Pummer (2017; forthcoming); Nebel (2018); and Li (unpublished manuscript).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([8]) Kraut (2018) contains a very interesting discussion of McTaggart's thesis, though he does not explicitly discuss the particular defenses of unrestricted intrapersonal aggregation that I do here. Nonetheless, Kraut's response to McTaggart and my own are, I think, relevantly similar.

([9]) For discussion of Mill's higher pleasures doctrine, see Brink (2013, chap. 3).

([10]) There is an enormously complex literature on sorites arguments and related phenomena involving vagueness (see, e.g., Hyde 2018 and Sorensen 2018). My appeal to indeterminacy has some affinities with multivalued approaches that reject the law of excluded middle. In the literature these approaches are sometimes associated with semantic treatments of vagueness, whereas I'm inclined to regard the vagueness in question as metaphysical, involving the scalar nature of normative properties. For discussion of metaphysical vagueness, see, e.g., Barnes (2010; 2014).

([11]) I think both sorites and transitivity arguments are at work in different parts of *Reasons and Persons*. Though Parfit defends the repugnant conclusion by a transitivity argument, some of his spectrum arguments about personal identity (1984, chap. 11) seem to be sorites arguments.

([12]) This is how Temkin (1996; 2012) and Rachels (1998; 2004) avoid the interpersonal repugnant conclusion. It is arguable that Parfit (2016) and Pummer (forthcoming) are committed to similar claims, for the discontinuities that they contemplate arguably require denying unrestricted transitivity.

([13]) This essay builds on but extends ideas in Brink (1993; 2006). I am grateful to Richard Arneson, Richard Kraut, Hon Lam Li, Doug Portmore, and Theron Pummer for valuable feedback on a draft.

David O. Brink

David O. Brink is Distinguished Professor of Philosophy at the University of California, San Diego. His research interests are in ethical theory, history of ethics, moral psychology, and jurisprudence. He is the author of Moral Realism and the Foundations of Ethics (New York: Cambridge University Press, 1989), Perfectionism and the Common Good: Themes in the Philosophy of T.H. Green (Oxford: Clarendon Press, 2003), Mill's Progressive Principles (Oxford: Clarendon Press, 2013), and Fair Opportunity and Responsibility (Oxford: Clarendon Press, 2021).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Consequentialism and Partiality

Diane Jeske

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.15

Abstract and Keywords

Consequentialism has often been criticized for its inability to accommodate the partiality toward intimates that most people regard as not only morally permissible but as morally required. Consequentialists have responded to this objection by attempting to show, in various ways, that such partiality can, in some sense, be justified by their theory. While the classical utilitarians such as Mill and Sidgwick claimed that adherence to rules of thumb advising partial behavior is a good strategy for maximizing value, in recent years, Peter Railton has defended what is known as indirect consequentialism. According to the indirect consequentialist, consequentialism is to be understood as a criterion of right action, not as a decision procedure for agents to employ in their practical reasoning. Thus, according to Railton and others, a good consequentialist agent will often act and be motivated in the partial manner supposedly advocated by common sense. I argue that consequentialist moves such as those taken by Railton et al. are misguided, because the real issue is not how much partial behavior a moral theory is able to justify but, rather, the way in which it justifies that partial behavior.

Keywords: agent-neutral, agent-relative, consequentialism, decision procedure, friendship, indirect consequentialism, intimacy, partiality

1. Introduction

THE philosophical literature aimed at undermining consequentialism as a moral theory is chock-full of thought experiments of the following form: you have the ability to save either person A or persons B and C, but you are not able to save all of A, B, and C. Consequentialism—at least in its classical maximizing versions—tells us that the right action is to save B and C rather than to save A. Barring some special demonstration that saving A will maximize value (by, for example, showing that A is likely to cure cancer if she lives, that B and C are serial killers, or that A is five years old and both B and C are over ninety years old), we should act to save the greater number. After all, generally two lives are more valuable (both intrinsically and instrumentally) than one, and we are generally justified in assuming this to be the case. So far, it seems, most of us will be on board with the

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

consequentialist. However, suppose that A is your mother (or child, or lover, or friend, etc.).[1] The consequentialist, it seems, must be indifferent to this fact about your relationship to A (at least considered in and of itself)[2] and continue to insist that you should save the greater number. But "common sense" balks at this claim, unwilling to deny that your special relationship to A, considered in and of itself, is relevant to the determination of what you ought to do. Thus, it seems, "common sense" is at odds with maximizing consequentialism.

(p. 239) These sorts of examples are meant to provide an objection to consequentialism in so far as it seems right to accept what John Cottingham calls "partialism," the doctrine that "it is (not merely psychologically understandable but) morally correct to favour one's own … [i.e.] those to whom the agent has some special relationship or personal tie."[3] Consequentialism, in its classical forms, is an impartial moral theory: it defines right action in terms of the maximization of intrinsic value regardless of the location of that intrinsic value. If I can maximize value by benefitting two strangers rather than by benefitting my friend, then, the consequentialist says, I ought to benefit the two strangers, even if the value thereby produced is only slightly more than that I would have produced by benefitting my friend. But many of us cannot help but think that choosing to benefit our friend is not only morally permissible but, in fact, morally required. This moral commitment to partiality is at least as deeply ingrained in us, it seems, as any moral commitment to making the world the best place that it can possibly be.

And many consequentialists have seemed to acknowledge this fact, as much of the consequentialist response to the objection from partiality is constituted by attempts to show that consequentialism can accommodate at least some version of partialism. The form that this attempt has taken has varied, but all forms have the aim of showing that, in some way, one can remain a good consequentialist agent while acting in the partial ways[4] supposedly supported by "common sense." Thus, the consequentialist argues, one can act as a friend ought to act—that is, by demonstrating partiality to one's friends—and still act as a consequentialist ought to act.

The success of these consequentialist attempts to accommodate partiality has been highly contested, with many critics claiming that, in the end, the partial behavior (and psychology) of a friend, has not been shown to be compatible with adherence to the consequentialist understanding of right action. In this chapter, I will present the main moves in this debate between consequentialists and their critics who claim that the former are unable to adequately respond to the objection from partiality. This is, in fact, a very large subject and so I cannot canvass all of the moves or even all of the responses to the moves that I do discuss. For example, I am not going to discuss moves to rule consequentialism[5] or to satisficing consequentialism,[6] instead focusing on the more traditional maximizing version(s) of act consequentialism.[7] I will leave open the questions as to whether we are concerned with actual consequences, probable consequences, expected consequences, (p. 240) and so on.[8] I will also leave it open as to what precisely has intrinsic value, talking vaguely about happiness or well-being. Generally I will be concerned with nonrelative

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

conceptions of intrinsic value, but will briefly show (in section 2) how a relativized conception avoids the objection from partiality.[9]

I think that the partiality debate reveals a fundamental rift between consequentialists and those moral philosophers such as myself who endorse partiality as a moral requirement. Attempts to show and to rebut the claim that consequentialists can think, be motivated, and/or act as friends ought to think, be motivated, and/or act are really just distractions from the fundamental issue of the moral status of special relationships considered in and of themselves. That consequentialism can justify partial motivation or behavior may serve to undermine intuitions that the theory is overly demanding, but it leaves open the question as to whether it justifies that partial motivation or behavior *in the right sort of way*. I argue that it fails to offer the correct justification and so is undermined regardless of the extent to which its approved agents walk and talk in just the way that some deontological agents walk and talk.

2. The Objection from Partiality

Before beginning I need to note that what I am calling the objection from partiality is only one of many possible versions of an objection from partiality. I am going to be talking about the moral legitimacy and status of partiality to our friends and other intimates.[10] But we could equally consider the moral legitimacy and status of partiality to those to whom we have made promises, to our fellow citizens, to our colleagues, to our allies in war, and so on. In cases other than those involving intimates, the nature of the special relationship and, thus, the relevant considerations will differ. However, the structure of the debate will often, in very general terms, be the same.

The objection from partiality, at bottom, is very simple. It begins from the supposedly common-sense intuition that morality not only allows but in fact requires us to be partial to our intimates. Consequentialism, it is claimed, cannot endorse such partiality. Thus, in so far as consequentialism often yields counterintuitive judgments in cases involving intimates, we ought to reject consequentialism.

(p. 241) Most of our lives, if they are to be morally justified, require the truth of the claim that morality at the least *allows* us to be partial to our intimates: parents buy toys, vacations, and pricey educations for their own children but not for other people's children; we offer support and encouragement to our friends but not to strangers or mere acquaintances; we take care of our elderly parents but not of other elderly people; and so on. Most of our other-regarding efforts are directed at those to whom we stand in some special relationship such as friendship or family. And we are partial to our own intimates even when there are people not only worse off than our intimates, but even when our children, friends, and parents are already doing fairly well and there are other people barely clinging on for dear life.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Given this sort of partial behavior, it initially seems to be obvious that it is not value maximizing. Consequentialism is a moral theory that understands right action in terms of the maximization of intrinsic value: right action is that action, out of all of the alternatives available to the agent, which maximizes net (actual, probable, expected) intrinsic value of the consequences (where the action may be considered as a constitutive consequence of itself). For example, suppose that the consequentialist is also a hedonist; that is, he or she accepts that all and only pleasure has intrinsic value. Then, according to this hedonistic consequentialism, the right action is that action which maximizes net pleasure (net pleasure being the sum of pleasure it produces minus the sum of the pain that it produces). Behavior exhibiting partiality, it seems, will often not maximize pleasure, especially when it is the behavior of those of us who are in relatively privileged positions: donating much larger sums to charity rather than taking our friends to the movies or buying video games for our children seems to be preferable if one is a consequentialist.

Thus, it seems that in order to be a good consequentialist agent, most of us would need to completely restructure our lives, altering our patterns of behavior almost entirely. And, of course, if we were to cease to demonstrate partiality to intimates, it is not at all clear that we could continue to sustain intimate relationships. Intimacy seems to require emotions and attitudes of special concern and also actions that express that special concern. But such actions are inevitably those that differentially benefit our intimates as opposed to other persons: one cannot express special concern for another if one does not devote greater efforts to promoting the well-being of the other than one does to promoting the well-being of all other persons.[11]

We can notice that the objection from partiality is only an objection to the classical versions of consequentialism in so far as those versions couple their maximizing principle of right action with a nonrelativized conception of intrinsic value. For the traditional consequentialist, if, for example, pleasure is intrinsically valuable, then it is intrinsically valuable simpliciter; that is, we do not need to specify *for whom* it is intrinsically valuable: the property of being valuable is not a relational property of pleasure. In particular, to say of some state of affairs that it is intrinsically valuable is not to say that some person or persons have some type of attitude such as approbation toward that state of affairs. If the consequentialist were to adopt a relativized conception of intrinsic value according **(p. 242)** to which what it is for something to have intrinsic value *for a person* is for it to be (intrinsically) desired or valued *by that person*, she can easily respond to the objection from partiality. After all, most people have very strong intrinsic desires for the welfare of their intimates and, at best, relatively weak intrinsic desires for the welfare of nonintimates. Thus, for most of us most of the time, the state of affairs produced by saving or benefitting our friend or parent has far more noninstrumental value for us than does the state of affairs produced by saving or benefitting some larger number of nonintimates. Thus, consequentialism coupled with a relativized conception of intrinsic value is going to have the same sorts of outcomes with respect to right actions as will the partialist, and thus it is not subject to the objection from partiality.[12] So from here on out I will be assuming a nonrelativized conception of intrinsic value.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

3. The Strategic Response to the Objection from Partiality

One sort of consequentialist response to the objection from partiality is a kind of nonresponse. The consequentialist might take the sort of route that, for example, Peter Singer takes in his oft-discussed "Famine, Affluence, and Morality" (1982). Singer essentially argues that at least those of us in privileged economic situations morally ought to completely change our way of life, that we cannot justify continuing to benefit ourselves and our intimates in the ways that we do when so many people are in genuinely desperate situations around the world. Singer argues that at least in our current empirical circumstances, we simply cannot morally justify our patterns of partiality toward ourselves and our intimates.

But most consequentialists do not bite the partiality bullet in the way that Singer does. John Stuart Mill and Henry Sidgwick both took what we can call the strategic response to the partiality objection: they argued that the best long-term strategy that individuals can adopt in their attempts to maximize value involves partiality to certain persons. In *Utilitarianism*, Mill says that "[t]he important rank, among human evils and wrongs, of the disappointment of expectation is shown in the fact that it constitutes the principal criminality of two such highly immoral acts as a breach of friendship and a breach of promise."[13] Mill is saying that failures in performing acts of friendship (supposedly including failures in the showing of differential concern) are wrong in virtue of their being instances of a broader class of actions, those actions which disappoint someone's expectations. Disappointing expectations is generally not value maximizing, and so one should follow the general rule of being partial toward one's friends. Similarly, in *The Methods of Ethics*, Sidgwick says that there are some services "which one is only willing to receive from genuine friends. It much promotes the general happiness that such services should (p. 243) be generally rendered. On this ground, as well as through the emotional pleasures which directly spring from it, we perceive Friendship to be an important means to the Utilitarian end."[14] Sidgwick also points out that our love for our intimates will lead us to put more effort into acquiring knowledge about how best to promote their interests, and so focusing our benevolent efforts on those close to us is likely to have better consequences than would a more indiscriminate charity.[15]

But, of course, this response to the objection from partiality has the same difficulties that any such consequentialist appeal to "rules of thumb" has: all such rules are necessarily defeasible if one is a consequentialist. The strategy relies on the empirical claim that, generally, apparently partial behavior has better consequences than explicitly impartial behavior has. A good consequentialist agent, it seems, must be ready to recognize times at which being partial to one's intimates is suboptimal and, thus, wrong. In simple cases like the one with which I began the paper—save one's intimate or two very similar strangers—it seems obvious that a good consequentialist ought to save the two strangers. And, thus, consequentialism remains at odds with what seems like the obviously correct answer: one is morally required, or at least permitted, to save one's intimate rather than

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

two strangers. In an attempt to get a stronger response to the objection from partiality, some consequentialists have abandoned the strategic response and adopted what has come to be known as indirect consequentialism.

4. Indirect Consequentialism

Most of the recent literature about consequentialism's ability to accommodate partiality has focused on the important distinction between a criterion of right action and a decision procedure. The consequentialist principle of right action holds that the right action is the action which, out of all of the alternatives available to the agent, produces the greatest net sum of intrinsic value for all persons affected in the long run. But consequentialists such as Peter Railton and David Brink have pointed out that this principle is properly understood as a *criterion* of right action: it states what it is for an action to be a right action. However, that criterion leaves open what kinds of cognitive and affective processes ought to precede action; in particular, it leaves open whether any consideration of the criterion itself ought to figure into the agent's thoughts. The *decision procedure* used by the utilitarian agent ought to be that decision procedure which is such that, in using it, the agent will produce the best consequences in the long run.

In making this point, Railton distinguishes between "subjective consequentialism" and "objective consequentialism": "[s]ubjective consequentialism … is a view that prescribes following a particular mode of deliberation in action; objective consequentialism … concerns the outcomes actually brought about, and thus deals with the question (p. 244) of deliberation only in terms of the tendencies of certain forms of decision making to promote appropriate outcomes."[16] If one is an objective consequentialist, one need not be committed to subjective consequentialism; in fact, an objective consequentialist can consistently claim that one ought *not* be a subjective consequentialist in so far as adopting such modes of deliberation might, in the long run, lead to less valuable outcomes than if one adopted nonconsequentialist modes of reasoning and decision-making.

So how does this appeal to subjective versus objective consequentialism provide the consequentialist with a response to the objection from partiality? Let us consider Mabel, who adopts the strategic approach discussed in the previous section: she generally focuses her benevolent efforts on her intimates, but keeps herself alert for cases in which deviating from her rule of being partial to her intimates will clearly not produce the best results, cases in which, for example, she can save ten children rather than her own one. Will Mabel, in adopting this strategy, be the best possible consequentialist agent? The suggestion made by indirect consequentialists such as Railton and Brink is that she might very well be a better consequentialist agent if she develops dispositions to aid her intimates even in cases in which doing so has worse results, that is, if she develops "dispositions that will sometimes lead [her] to violate [her] own criterion of right action."[17] By doing so, Mabel will be taking the attitude of commitment to her loved ones that, it is claimed, is necessary in order to sustain the great good of intimate relationships and to avoid the alienation produced by viewing all of her actions and relationships from the ob-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

jective consequentialist perspective. Thus, both the partial behavior and the partial attitudes advocated by “common sense” can also be advocated by the sophisticated consequentialist, that is, the consequentialist who recognizes that being explicitly consequentialist in one’s deliberations and attitudes may undermine the pursuit of optimal outcomes.[18]

Railton is claiming more than just that the actions of the sophisticated consequentialist can mimic those of the partialist. He is also claiming that sophisticated consequentialists such as Mabel will have attitudes, emotions, and commitments that look very much like those had by partialists. This is important because many advocates of the objection from partiality have claimed that being a good friend is a matter not only of acting in the right sort of way but also of deliberating and feeling in the right way. The subjective consequentialist has as her overriding goal the promotion of impersonal (i.e., nonrelativized) value, and, thus, it seems, she subordinates all of her other goals, such as the well-being of friends and her relationships with those friends, to the achievement of maximal net value. It seems, then, that the subjective consequentialist must view her intimates and her relationships with them from an objective, impartial viewpoint and, thus, will inevitably suffer from “alienation” from her intimates.

(p. 245) Whatever the force of this objection to the impersonal nature of subjective consequentialism, Railton and Brink insist that it poses no problem for the objective or sophisticated consequentialist. Objective consequentialists will cultivate dispositions to act in accordance with rules such that, by cultivating such dispositions, they will best promote overall value in the long run, and they will view these rules not as mere rules of thumb (as the advocate of the strategic response does): the rules “will not function as aids in utilitarian deliberation. Rather, moral rules [such as “show differential concern for your intimates”], on this view, should be appealed to and applied more or less strictly and uncritically in most cases.”[19] On a day-to-day basis, the dispositions guiding the objective consequentialist will look much the same as those that guide the partialist.[20] So if both the objective consequentialist and the partialist will save A who is her intimate rather than B and C who are not her intimates and have stable dispositions to behave in such a way, then has the objection from partiality been diffused?

5. Consequentialism and Intimacy

Many partialists have remained dissatisfied with the moves made by the indirect consequentialist in response to the objection from partiality. One source of dissatisfaction is the supposed attitudes which the consequentialist must have with respect to her intimates, attitudes, it is argued, that are not compatible with true intimacy.

One sort of objection to consequentialism can, I believe, be dealt with fairly quickly. Neera Badhwar Kapur claims that the consequentialists must accept “an instrumental justification of friendship” and that this type of justification “is logically inconsistent with the attitudes and motivations” of the true friend.[21] She claims that you ought to think of your friend, “I place a special value on you out of friendship and not out of consequential-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ist considerations," but, if you are a consequentialist, you must add, "but as a consequentialist agent I do so only so long as, all things considered, valuing you thus promotes the overall good."[22] Kapur is claiming that someone who is a subjective consequentialist cannot value her friends as she ought, because she cannot differentially or partially value some persons independent of the contribution of those persons and of her relationship with those persons to the overall good.

This, however, is confused. The overall good is just the sum of all states of affairs with intrinsic value. It is true that for any such intrinsically valuable state of affairs that it is a means to the greater good, but it is a constitutive means, not a causal means. It is not that it is valuable in virtue of its contribution to the greater good; rather, its value is contributory to the sum of valuable states of affairs that is the overall good. So if my (p. 246) friend's welfare and my intimate relationship with her are both intrinsically valuable to a certain degree, they continue to have that degree of value regardless of whether I ought to sacrifice them in my pursuit of maximal value. And I can continue to subjectively care more about my friend and my relationship with her even if I realize that I am morally required to make a choice contrary to the well-being of my friend or of my relationship with her. In such a case, my love for my friend will be reflected in my subsequent attitudes of sorrow and regret.

But, it can be pointed out, sorrow and regret are types of pain and so are plausibly understood as bad. At some point, I need to consider, even if I am an indirect consequentialist, whether altering my attitudes and dispositions would produce better consequences than would continuing my partially disposed status quo. As Railton says, the indirect consequentialist's "motivating structure meets a counterfactual condition: while he ordinarily does not do what he does simply for the sake of doing what's right, he would seek to lead a different sort of life if he did not think that his were morally defensible."[23] The indirect consequentialist is still, after all, a consequentialist: her disposition to be partial is justified in so far as it brings about the best consequences in the long run. If and when that disposition ceases to do so, it ceases to be justified and the sophisticated consequentialist must become a subjective consequentialist.

Dean Cocking and Justin Oakley have argued that because the indirect consequentialist must be governed by this counterfactual condition shows that an appeal to indirect consequentialism does not provide an adequate response to the objection from partiality. They define a regulative ideal as an "internalized disposition to direct one's actions in certain ways."[24] The sophisticated consequentialist, as Railton acknowledges, has as a regulative ideal a disposition to enter into, maintain, and terminate intimate relationships in accordance with the demands of consequentialism: the acceptance and terminating conditions (governing conditions) of intimate relationships for the indirect consequentialist are the same as those held by the direct consequentialist and are to be understood solely in terms of a particular relationship's maximization of value in the long run. The only difference between the direct and the indirect consequentialist is the circumstances under which those governing conditions will be in the front of her mind.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Cocking and Oakley insist that the consequentialist governing conditions on intimate relationships are not compatible with the common-sense understanding of the partiality required for friendship—they insist that friends must be partial to their relationship with one another such that the mere fact that more value could be promoted by abandoning the relationship does not motivate them to actually abandon the relationship. They claim (p. 247) that "[t]rue and good friends … will have a motivational disposition which involves a preparedness to act for the friend, such that the claims of friendship will sometimes trump the maximization of agent-neutral value."[25]

Responding to Cocking and Oakley in defense of the indirect consequentialist's reply to the argument from partiality, Elinor Mason claims that they have misread Railton's counterfactual condition. Cocking and Oakley claim that the indirect consequentialist must have as one of her governing conditions for each intimate relationship that she will terminate it when it ceases to be value-maximizing. In other words, Cocking and Oakley understand the indirect consequentialist as refraining from evaluating each *act of friendship* in consequentialist terms but as still evaluating each *friendship* in its entirety in consequentialist terms (although not constantly or always explicitly). But, Mason insists, the indirect consequentialist does not have such a governing condition on particular friendships but only on her general pro-friendship disposition.[26] To have such a pro-friendship disposition, in fact, she insists, is to be committed to particular friendships in just the way that Cocking and Oakley insist genuine friends ought to be committed to their friendships; that is, sophisticated consequentialists with a pro-friendship disposition should view the fact of a friendship as able to trump considerations of the maximization of value.[27]

An important feature of any defense of consequentialism in response to the objection from partiality that takes the route of Railton, Brink, and others is that, ultimately, it depends upon an appeal to purported empirical truths about the causal relationship between forms of deliberation, characters, or ways of life and the production of maximal value. This is hardly surprising, given that consequentialists have only one fundamental moral principle—the principle that connects right action to action productive of overall best consequences—and any other moral principles have to be derivative from that fundamental principle in conjunction with empirical truths about how to bring about various states of affairs. So if deliberating in other than an explicitly consequentialist manner, having a pro-friendship disposition, and being partial to one's intimates on a daily basis are to be justified, then the consequentialist must supply empirical premises that support the conclusion that so deliberating, having such a disposition, and being partial are causally productive of better consequences than deliberating otherwise, having a different disposition, or being impartial on a daily basis.

But philosophers are not in the business of testing empirical causal hypotheses. More importantly, it is unclear how we could test the empirical claims that play a role in the arguments of Railton et al. How could I possibly determine if my pro-friendship and other partial dispositions have better consequences than would alternative dispositions? In any such process of testing, I would most likely lose any opportunity to return to my current dispositional structure. Further, my having an anti-friendship disposition in a world of

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

people who continue to have pro-friendship dispositions is likely to have very (p. 248) different consequences than would my having such a disposition in a generally anti-friendship disposition world.

However, the real issue is not whether the indirect consequentialists are correct in their empirical hypotheses. The real issue is the moral or rational status of the fact that I stand in an intimate relationship with another person, and that is what we now need to consider.

6. The Moral and Rational Significance of Intimacy

Consider the following remarks by Scott Woodcock:

> But the question of when consequentialists ought to revert back to direct methods of promoting the good at the expense of their friendships is important, for even if the in-principle objection to incorporating friendship within consequentialism can be avoided, a practical objection looms if empirical circumstances are such that progressive versions of consequentialism still end up dissolving friendships in non-ideal contexts.[footnote omitted] For many, it is seriously counterintuitive for friendship to be precluded by the practical application of an ethical theory in current circumstances regardless of whether the theory is consistent with friendship as a matter of principle.[28]

I think that Woodcock is entirely right that for many people the crucial issue is the *extent* to which each of us is morally justified in acting in a partial manner with respect to her friends. But, in spite of being a strong advocate of partiality as being justified by intimacy, I do not think that this is the real issue. I am not convinced that, in fact, my balance of reasons in my current circumstances supports the extent to which I am partial to my intimates, and that is not because I believe myself to display more partiality than do others in my socioeconomic circumstances; rather, I think it is because most of us have strong desires to act in accordance with the reasons generated by intimacy but at most weak desires to act on our reasons to promote maximal welfare. Thus, we rather naturally give more weight to our reasons to be partial than we do to our reasons to be impartial.[29] But, as a matter of fact, continuing our current patterns of partiality is likely, I am inclined to think, to continue to widen the gap between the rich and the poor, thereby making the poor even worse off, thereby weakening the relative strength of our justification of being partial to our intimates, no matter what moral theory we hold.[30]

(p. 249) The real crux of the debate is about the status of our reasons to be partial to our intimates. Consequentialists, be they of the subjective or of the objective variety, are committed to the view that our reasons to be partial to our intimates are only derivatively agent-relative reasons. For consequentialists, our fundamental reasons are all agent-neutral; that is, they are reasons that make no essential reference to the agent whose rea-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

sons they are: each of us has a fundamental reason to promote maximal nonrelativized value. In so far as I have a reason to focus my efforts on my intimates it is derivative from that fundamental reason—I have such a reason in so far as and only as far as doing so is a means to promoting maximal value.[31] Many of us who are partialists, however, think that we have fundamental agent-relative reasons, that is, reasons that make ineliminable reference to the agent whose reasons they are. I have reason to be partial to my intimates in virtue of their being my intimates, not because being partial to them is the best way for me to promote value—this is a claim that all consequentialists, sophisticated or unsophisticated, must reject.

It is important to see that commitment to fundamental agent-relative reasons of partiality does not entail any sort of commitment to how often in our daily lives, given the way that the world is currently structured and given our own socioeconomic positions, we are allowed to demonstrate partiality to our intimates.[32] If partialists think that we have agent-neutral reasons to promote intrinsic value (or agent-neutral reasons of some other kind) in addition to our agent-relative reasons of partiality, then they should allow that our agent-neutral reasons can trump our agent-relative reasons. It then becomes an open question as to how the competition between the two types of reasons plays out in any given life.

Some partialists will insist that a good life requires intimate relationships, intimate relationships require partiality, and, thus, there must be at least agent-relative (or agent-centered) permissions to be partial to our intimates.[33] There are, I think, at least two worries about such an appeal to agent-relative permissions. First, and, I think, most importantly for the partialist, they do not constitute any sort of requirement that one be partial to one's intimates, only an allowance: one will do no wrong in certain cases when one acts to benefit an intimate instead of promoting maximal well-being. However, it will also be the case that one will do no wrong in aiming at maximal well-being instead if consequentialism is supplemented only with permissions. Further, any sensible account of such permissions must allow that the area of moral freedom for the agent which they carve out will become smaller as the claims of impersonal welfare become more urgent in one way or another. Thus, their success at accommodating various "common-sense" (p. 250) intuitions will depend, as did the appeal to indirect consequentialism, on various contingent, empirical truths about the world and the agent's epistemic and causal position within the world. I am inclined to think that any plausible theory of practical rationality *should* have that result.

What matters, then, in adjudicating between the claims of the consequentialist and of the partialist is not whether it is possible for the consequentialist to accommodate a life that mimics that of a partialist, because what a partialist life looks like will depend on various contingent features of her life. The question for the moral theorist has to be what the correct account of reasons for action is in cases such as the one with which I began this paper: you can save A or you can save B and C but you cannot save all of A, B, and C. A is your closest friend in the world and there are no other relevant differences that would make a difference to the consequences (such as one being a cancer researcher nearing

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

success in her search for a cure, one or more being serial killers or genocidal maniacs, or one or more being aged homeless people who will not be missed by anyone). In such a case, the consequentialist, be she subjective or objective, sophisticated or unsophisticated, must say that the right action is to save B and C. A sophisticated consequentialist will, in fact, save her close friend A, but, as Troy Jollimore points out, such actions "are neither admirable nor even permissible, but are cases of "blameless wrongdoing." Such a view seems to provide an excuse rather than a justification" for partial acts of friendship.[34] This is because there are no fundamental agent-relative reasons for you to save your friend, only reasons derived from the agent-neutral reason to promote nonrelativized value. The fact of your friendship with A provides no independent consideration for you to factor into your practical deliberations.[35]

In *The Right and the Good*, W. D. Ross criticized utilitarianism on the grounds that it fails to grant an act of promise-keeping any independent moral significance. He says of the "plain man" that "[w]hat makes him think it right to act in a certain way is the fact that he has promised to do so—that, and, usually, nothing more."[36] Of course, Ross himself views the fact of the promise as making it prima facie right to act in the way that would constitute keeping the promise, and so deliberation, at least at some level, requires consideration of other factors in order to determine if so acting would be an all-things-considered right action.[37] But Ross's central point could also be made, by the partialist, about the case of A, B, and C: "what makes it at least prima facie right to save one's close friend A is the fact that A is your close friend." And, if the value of the consequences of saving A were the same as the value of the consequences of saving B and C, the consequentialist still must deny that you ought to save your friend. We have to consider these sorts of isolated thought experiments to get to the crux of the partialist position and to see why various sophisticated moves on the part of the consequentialist simply do not get to that central partialist claim.

(p. 251) There are, of course, more moves for the consequentialist to make. One possibility is to say that our partialist intuitions result from our being habituated to consequentially justified ways of life and of deliberation, thereby offering a type of error theory to explain away what appear to be deep commitments to partiality. Or the consequentialist can try to say that we are covertly dragging into consideration further good consequences of saving our friend that would usually be present in the real world as opposed to an artificially isolated hypothetical situation. At some point we just have to decide whether our commitment to consequentialist intuitions is strong enough to outweigh our commitment to partialist intuitions. I myself am inclined to think that for even many consequentialists the answer is "no," given the great efforts they make to justify at least apparent partialism.[38]

References

Badhwar Kapur, N. 1991. "Why It Is Wrong to Be Always Guided by the Best: Consequentialism and Friendship." *Ethics* 101, no. 3: 483–504.

Baron, M. 1991. "Impartiality and Friendship." *Ethics* 101, no. 4: 836–857.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Brink, D. 1986. “Utilitarian Morality and the Personal Point of View.” *The Journal of Philosophy* 83, no. 8: 417–438.

Cocking, D., and Oakley, J. 1995. “Indirect Consequentialism, Friendship, and the Problem of Alienation.” *Ethics* 106, no. 1: 86–111.

Cottingham, J. 1986. “Partiality, Favouritism, and Morality.” *The Philosophical Quarterly* 36, no. 144: 357–373.

Driver, J. 2005. “Consequentialism and Feminist Ethics.” *Hypatia* 20, no. 4: 183–199.

Jackson, F. 1991. “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection.” *Ethics* 101, no. 3: 461–482.

Jeske, D. 2008. *Rationality and Moral Theory: How Intimacy Generates Reasons*. New York: Routledge.

Jeske, D., and Fumerton, R. 1997. “Relatives and Relativism.” *Philosophical Studies* 87, no. 2: 143–157.

Jollimore, T. 2000. “Friendship Without Partiality?” *Ratio* 13, no. 1: 69–82.

Mason, E. 1998. “Can an Indirect Consequentialist Be a Real Friend?” *Ethics* 108, no. 2: 386–393.

Mill, J. S. 1979. *Utilitarianism*. Indianapolis: Hackett. (Original work published 1861.)

Norcross, A. 1997. “Consequentialism and Commitment.” *Pacific Philosophical Quarterly* 78, no. 4: 380–403.

Pettit, P. 1994. “Consequentialism and Moral Psychology.” *International Journal of Philosophical Studies* 2, no. 1: 1–17.

Powers, M. 2000. “Rule Consequentialism and the Value of Friendship.” In *Morality, Rules, and Consequences: A Critical Reader*, edited by B. Hooker, E. Mason, and D. E. Miller, 239–254. Edinburgh: Edinburgh University Press.

Railton, P. 1984. “Alienation, Consequentialism, and the Demands of Morality.” *Philosophy and Public Affairs* 13, no. 2: 134–171.

(p. 252) Ross, W. D. 2002. *The Right and the Good*. Edited by P. Stratton-Lake. Oxford: The Clarendon Press. (Original work published 1930.)

Scheffler, S. 1982. *The Rejection of Consequentialism*. New York: Oxford University Press.

Sidgwick, H. 1981. *The Methods of Ethics*. Indianapolis: Hackett. (Original work published 1907.)

Singer, P. 1972. “Famine, Affluence, and Morality.” *Philosophy and Public Affairs* 1, no. 3: 229–243.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Slote, M. 1985. *Common-Sense Morality and Consequentialism.* London: Routledge and Kegan Paul.

Wilcox, W. H. 1987. “Egoists, Consequentialists, and Their Friends.” *Philosophy and Public Affairs* 16, no. 1, 73–84.

Williams, B. 1981. “Persons, Character, and Morality.” In *Moral Luck*, 1–19. New York: Cambridge University Press.

Woodcock, S. 2010. “When Will Your Consequentialist Friend Abandon You for the Greater Good?” *Journal of Ethics and Social Philosophy* 4, no. 2: 1–23.

Notes:

([1]) From now on, I will use “friends” or “intimates” as catch-all terms for persons who stand in intimate relationships to us.

([2]) In other words, considered independently of the consequences that such a relationship might produce. Perhaps your guilt at failing to save your mother is so intense that you will be unable to function—the negative value of such a result would, of course, be relevant to determining how you ought to act in the circumstances.

([3]) Cottingham (1986), 357–358.

([4]) When I use locutions such as “partial behavior” or “partial action,” I am using them as shorthand for behavior or actions that demonstrate partiality toward intimates. I, of course, am not talking about behavior or action that is somehow incomplete or cut short.

([5]) For a rule-consequentialist defense of the partiality of friendship, see Powers (2000). See also Hooker, Chapter 23, this volume.

([6]) See Slote (1985) and Chappell, Chapter 26, this volume.

([7]) Act consequentialists evaluate actions as right or wrong in virtue of whether those actions themselves are value maximizing. Rule consequentialists, on the other hand, evaluate acts as right or wrong in virtue of whether those acts are in accordance with the appropriate rule, e.g., a rule such that if most people usually followed the rule, then value would be maximized.

([8]) For a response to the objection to partiality that appeals to a version of consequentialism defined in terms of the consequences that are probable from the agent’s epistemic perspective, see Jackson (1991). See also Jackson, Chapter 17, this volume.

([9]) See also Hammerton, Chapter 3, this volume; and Hurley, Chapter 2, this volume.

([10]) Work needs to be done in order to specify the nature of the relationship between two people that is necessary and sufficient to render them intimates. However, I am going to proceed taking friendship and at least close family relationships to be our paradigms of

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

intimate relationships. Nothing in what follows depends upon a more precise statement of the nature of the special relationship of intimacy. For an account of the nature of intimacy, see Jeske (2008).

([11]) See Jeske (2008) and Jollimore (2000), 72–73, 77.

([12]) See Jeske and Fumerton (1997).

([13]) Mill (1979), 59.

([14]) Sidgwick (1981), 434.

([15]) Sidgwick (1981), 431–432.

([16]) Railton (1984), 152. This terminology of subjective versus objective consequentialism can be misleading if one is thinking about the distinction between subjective and objective value. We just have to be careful to keep Railton's use of the terms in mind.

([17]) Railton (1984), 157.

([18]) Railton (1984), 153.

([19]) Brink (1986), 425.

([20]) See Baron (1991), 842, 844.

([21]) Badhwar Kapur (1991) 488. For a discussion and rejection of this sort of claim, see also Pettit (1994).

([22]) Badhwar Kapur (1991), 493.

([23]) Railton (1984), 151. There is, of course, a more radical sort of consequentialist move: the consequentialist agent might decide that she will be a better consequentialist agent if she makes herself into a deontologist. If she is successful in her endeavor to change her beliefs, then she will later regard her partiality as justified on deontological grounds, although her earlier self (or a consequentialist observer) would say that both her partiality and her deontological understanding of its justification are justified on consequentialist grounds. This would be similar to how someone convinced by Pascal's Wager would understand her transition from nonbeliever to believer.

([24]) Cocking and Oakley (1995), 90. For a similar discussion, see Wilcox (1987), 78–79.

([25]) Cocking and Oakley (1995), 109.

([26]) Mason (1998), 390.

([27]) Mason (1998), 391. For another defense of a consequentialist's ability to be committed to her friends, see Norcross (1997).

([28]) Woodcock (2010), 3.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([29]) There is a further issue as to whether we have reasons grounded by our desires. If we do, the extent of our partiality may in fact be justified.

([30]) Of course, this is just so much armchair empirical hypothesizing on my part and so is to be taken with quite a few grains of salt.

([31]) See also Driver (2005), 193.

([32]) If intimacy requires past interactions that demonstrate concern, then an intimate relationship presupposes past actions of partial treatment. (See Jeske [2008].) Whether or not one was justified in the past in being partial, I think that the fact that one is now intimate with someone grounds reasons for continued partiality.

([33]) See Scheffler (1982) for the original discussion of such permissions. See also Brink (1986) for the suggestion that concerns such as that of the partialist are really "worries about the justification or supremacy of moral demands, not about the correctness of a utilitarian account of morality" (433).

([34]) Jollimore (2000), 70.

([35]) See Woodard, Chapter 9, this volume.

([36]) Ross (2002), 17.

([37]) So Ross would not agree with Williams that if a rational deliberator goes on to think, "And I ought to keep my promise, all-things-considered, including what good could be brought about by my breaking my promise," has "one thought too many."

([38]) I would like to thank Richard Fumerton and Douglas Portmore for comments on an earlier version of this chapter.

Diane Jeske

Diane Jeske is Professor of Philosophy at the University of Iowa, where she has taught since 1992. Her work has focused on the nature and significance of intimate relationships and how that significance ought to be reflected in moral theory. She is the author of Rationality and Moral Theory: How Intimacy Generates Reasons (Routledge, 2008), The Evil Within: Why We Need Moral Philosophy (Oxford University Press, 2018), and Friendship and Social Media: A Philosophical Exploration (Routledge, 2019).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Consequentialism and Promises

Alida Liberman

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.16

Abstract and Keywords

I explore the debate about whether consequentialist theories can adequately accommodate the moral force of promissory obligation. I outline a straightforward act consequentialist account grounded in the value of satisfying expectations, and I raise and assess three objections to this account: that it counterintuitively predicts that certain promises should be broken when common-sense morality insists that they should be kept, that the account is circular, and Michael Cholbi's argument that this account problematically implies that promise-making is frequently obligatory. I then discuss alternative act consequentialist accounts, including Philip Pettit's suggestion that promise-keeping is an intrinsic good and Michael Smith's agent-relative account. I outline Brad Hooker's rule consequentialist account of promissory obligation and raise a challenge for it. I conclude that appeals to intuitions about cases will not settle the dispute, and that consequentialists and their critics must instead engage in substantive debate about the nature and stringency of promissory obligation.

Keywords: promises, promissory obligation, promise-making, promise-keeping, utilitarianism, counterexamples, intuitions, pro tanto obligations, act consequentialism, rule consequentialism

1. Introduction: Framing the Debate

IT is widely assumed by both philosophers and ordinary folks that we are generally morally obligated to keep our promises. A perennial worry about consequentialist moral theories is that they cannot account for the seemingly strict obligation to keep our promises in a wide range of cases, including those in which better overall consequences would result from promise-breaking than from promise-keeping. Can consequentialism adequately explain and accommodate the moral force of promissory obligation? And what impact does the answer to this question have on the plausibility of consequentialism as a moral theory?

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Consequentialists (and their critics) will obviously care about the answers to these questions. But other ethicists have at least two reasons to be interested in this debate as well. First, assessing the relationship between consequentialism and promises sheds light on our understanding of promissory obligation more broadly. And second, it informs what approach we should take to moral theorizing from imagined cases.

In this essay, I argue that the accommodation of promissory obligation raises serious challenges for consequentialist views of various kinds, but that this is simply one strike against them, which does not by itself entail that the views are implausible. In section 2, I explain straightforward act consequentialist accounts of promising. I discuss three important challenges for these accounts in section 3. I introduce and assess alternative act consequentialist theories in section 4. I outline rule consequentialist theories in section 5 and assess them in section 6, before closing with a brief discussion of the upshots of this debate in section 7.

(p. 290) 2. Straightforward Act Consequentialist Accounts

Straightforward act consequentialists determine the morality of actions by assessing whether they result in the best overall consequences.[1] On such accounts, promissory obligation stems from the fact that promise-making raises the promisee's expectations that the promisor will act as she has pledged to.[2] On these views, we are morally obligated to keep our promises because, and only to the extent that, fulfilling these expectations leads to better results than does failing to fulfill them. There are a number of ways in which promise-keeping can increase happiness or well-being. Most obviously, the promisee generally wants the promised action to occur and is made better off if the promise is kept than if it is broken. This is especially so if the promisee has (perhaps detrimentally) relied on the promise in her planning—say, if A has promised to drive B to the airport and B has failed to make alternate transportation arrangements. As Mill puts it, "few hurts which human beings can sustain are greater, and none wound more, than when that on which they habitually and with full assurance relied, fails them in the hour of need; and few wrongs are greater than this mere withholding of good; none excite more resentment, either in the person suffering, or in a sympathising spectator" (1863, 60–61).

Act consequentialists point out that, in many cases, *promising* to perform a desired action has greater utility than does simply stating your intention to perform it, or performing it without first articulating any plans to do so. This is in part because promising makes performance more likely than it would otherwise be, since promisors risk social sanctions if they do not follow through. Hume takes this observation a step farther and argues that promise-making is essential for attaining mutually beneficial reciprocal exchanges with others—such as your helping me with my harvest this week in exchange for my helping you with yours next week—that would not be possible without a robust convention of promising.[3] Individual acts of promise-keeping support and maintain (p. 291) this conven-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

tion, and thereby increase everyone's well-being. The formation of expectations can also be valuable in itself, insofar as this eases the promisee's mind about whether the action will occur. As Jan Narveson notes, the promisee receives increased utility not just from the performance of the promised action but also from "being able to look forward to it, to plan on it, to adjust your activities in such a way as to make them harmonize with the to-be-enjoyed activity" (1971, 216).

Act consequentialists argue that grounding promissory obligation in expectations gives us a good explanation (and, in some cases, a better explanation than deontological accounts can offer) of some of the core features of our promising practice. Although P. S. Atiyah does not endorse an act utilitarian account of promising, in discussing such accounts he argues that the utilitarian value of satisfying expectations can explain all of the following: (1) why promises must be communicated to be binding; (2) why promises must be accepted to be binding; (3) why promises vary in strength on the basis of how robust the raised expectations are;[4] (4) why promises should sometimes not be kept (e.g., to avoid moral disaster); and (5) why promises stemming from fraud, coercion, misunderstanding, and the like do not create moral obligations (1981, 45–48). Failing to respect any of these constraints would make our promises less useful or beneficial than they could otherwise be.

3. Criticisms of Straightforward Act Consequentialist Accounts

3.1. Counterexample Cases of Inappropriate Promise-Breaking

An important desideratum for a theory of promissory obligation is that it captures the right set of cases. The most central and enduring objection to act consequentialist accounts of promising is that they entail that agents are morally permitted or even obligated to break their promises in cases in which it intuitively seems that the promises should be kept. An influential version of this criticism comes from W. D. Ross (1930, 34–39). Ross invites us to imagine a promise whose keeping creates 1,000 units of good for A (to whom the promise was made) and whose breaking creates 1,001 units of good for B. (p. 292) (As he notes, this goodness calculation will have to be nuanced enough to include not only the promisee's disappointed expectations, but also the extent to which the broken promise weakens our valuable social convention of promising or leads the promisor to more readily break promises in the future, both of which will lead to a loss of net utility.) Utilitarians (at least, of the sort Ross is criticizing) argue that it is self-evident that the right thing to do is whatever leads to the best consequences. But Ross points out that this is not true in this case: to the contrary, common-sense morality holds that a much greater difference in benefit is necessary to justify promise-breaking, and that the promise to A should be kept even though breaking it would bring about greater benefit.[5]

More generally, it has been frequently objected that consequentialist theories that ground promissory obligation in the value of satisfying the promisee's expectations cannot explain why we are morally obligated to keep promises even in cases in which no bad conse-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

quences would result from promise-breaking. These include what David Owens calls "bare wrongings," in which breaking the promise would not lead to harm and keeping the promise would not lead to benefit. Other examples include promises made in secret where the promisee (and others) will never know whether the promise is kept, as well as promises made to people on their deathbeds.[6] Ross (1939, 104–105) offers a case of this sort: B promises a dying A that he will give A's property to C. But C knows nothing about the promise, and D would make better use of the property. Utilitarianism demands that B give the property to D, but Ross claims that "this utilitarian way of considering such a case is not the way in which honest men actually would consider it" (105). Rather, "most thinking people" know "that there is a *prima facie* duty to fulfil promises, distinct from the *prima facie* duty to produce what is good" (105).[7]

3.1.1. Consequentialist Responses

Some act consequentialists happily grant the existence of cases in which promise-keeping is not required on consequentialist grounds. For example, Peter Singer argues that there is no need for an absolutist rule requiring promise-keeping; rather, "all that is necessary is that there be habits of telling the truth and keeping promises unless there is a clear disutility in doing so which outweighs the benefits of preserving the useful practices and fulfilling the expectations aroused" (1972, 102). And Jan Narveson claims that most people would not bother to keep a promise if it were clear that the promisee no (p. 293) longer cared about their doing so, claiming that they "do not believe that their obligation is independent of the actual good they are doing the promisee, and actual disappointment that would be caused by their non-performance" (1971, 220).

Other act consequentialists argue that the problem is not as stark as their critics make it out to be, and that promise-keeping will be required in almost every case for purely consequentialist reasons. As Sidgwick puts it, "the importance to mankind of being able to rely on each other's actions is so great, that in ordinary cases of absolutely definite engagements there is scarcely any advantage that can counterbalance the harm done by violating them," although he grants that promises should be broken in certain sorts of unusual cases (*ME* book 4, chap. 3, sec. 4). Such claims doubt whether there will ever be real-life cases in which more good is done by breaking a promise, but it is obvious to all that the promise ought to be kept anyway.

Some act consequentialists attempt to explain away our contrary intuitions in other ways. For example, Alistair Norcross (2011) outlines a case in which John promises his dying grandfather that he will return a statue to the long-abandoned temple from which the grandfather stole it years ago. On his way to return the statue—which is worth little—a merchant offers John a moderate sum of money for it. Assume that it would maximize utility for John to sell the statue. Norcross grants that we are not inclined to blame John for keeping his promise, and we may be inclined to think that this is what he should do. But he offers two explanations of why this may be the case. First, it could be that "as a matter of psychological fact, it would be difficult for someone with a morally appropriate commitment to keeping promises to break one in such circumstances," even though this would in fact be the right thing to do (2011, 232–233). While Norcross is not explicit about this, he

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

seems to be presuming that a person of good character would be unable or unlikely to break a promise in John's situation. Perhaps our intuition that John ought to keep his promise is being confused with the intuition that promise-keeping would exhibit virtue or would reflect well on his character.

Second, it could be that John's general disposition to keep promises to his grandfather is utility-maximizing, such that had he "been able on this occasion to perform the objectively right action, he would have been less committed to keeping his promises. If he had been less committed to keeping his promises, perhaps he would have done less good in the long run" (231). Norcross argues that it is possible for John to remain committed to objective consequentialism (or acting in ways that actually maximize utility) while adhering to a nonconsequentialist decision procedure.[8] This possibility does not refute the alleged counterexamples raised by the critic; it remains the case that John's promise-keeping is *in fact* wrong, and this is precisely what is up for dispute. However, if it were (p. 294) true that John would be able to successfully keep his promises in general (and thereby maximize utility) *only if* his habit of promise-keeping were so robust that it led him to keep a non-utility-maximizing promise to his grandfather, keeping this promise would be justified on utilitarian grounds. While this may be true for some cases, I am skeptical that this explanation will work across the board.

3.1.2. Are Our Intuitions about Cases Reliable?

A major challenge for thinking about this line of criticism is that the counterexample cases raised by critics (i.e., non-utility-maximizing promises that seem morally obligatory to keep) are usually extremely abstract. It is doubtful that our intuitions about such abstract cases are very reliable. But it is not easy to come up with specific cases that clearly get the postulated results.[9] If the additional good accrued to the promise-breaking option is very tiny, it will not be obvious that it causes more good to break it after all, and it will accordingly not be obvious that the promise should be broken on consequentialist grounds. But if the additional good is very large, it may seem obvious to even nonconsequentialists that we should break the promise. As Norcross says, "it is difficult to describe a case in detail, such that it is obvious both that breaking the promise has better consequences than keeping it and that it would be wrong to break it" (2011, 220). These concerns are exacerbated by uncertainty: it is often unclear whether breaking or keeping a promise will do the most good in a particular case. And promisors themselves—with their human tendencies toward self-interest—are generally not good judges of this. When faced with uncertainty about particular cases, the best option may be to default to rules of thumb that are more generally reliable. As Atiyah puts it, "there are so many circumstances in which it is in practice impossible to be sure what is best on the whole that rules like the rule that promises should be kept, become almost conclusive [on utilitarian grounds] over a wide spectrum of activity" (1981, 77).[10]

This highlights an important theoretical point: if we are going to appeal to intuitions about cases in assessing the plausibility of a moral theory, we should do our best to ensure that the cases are as concrete and realistic as possible. Otherwise, we risk generating intuitions that are not reliably tracking our genuine moral reactions to cases, but are

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

simply reflecting our existing theoretical commitments or building in our own idiosyncratic presumptions about uncertain outcomes. I am more confident than is Norcross about whether we can construct concrete cases that generate the critic's intuition. For example, suppose Alyssa and Brendan are strangers sitting next to each other (p. 295) on an airplane that has been stranded on the tarmac for several hours. An exasperated Brendan asks Alyssa to promise him that she will write a one-star online review of the airline when she gets off of the flight. Alyssa makes the promise. Suppose that no negative consequences will result from her failing to write the review: Brendan will never find out about it and so cannot be disappointed, no one else overheard her make a promise, Alyssa is already so inclined to break casual promises that breaking yet another will not affect her future reliability, and one more negative review will not impact the airline's overall rating or bottom line. Assume also that Alyssa has no independent moral obligation to keep this promise.[11] Writing the review will slightly decrease net utility: while it will be easy, it will take a few minutes, and will cause Alyssa to dwell momentarily on her frustrations. It nevertheless seems to me that Alyssa has at least a weak moral obligation to write a review, simply in virtue of having promised to do so. Act consequentialist theories cannot explain the existence of a moral obligation in cases like this.

However, I grant that our intuitions about such cases might be murky: while we may be confident in our judgements about whether promises ought to be kept when much is at stake, we may not have strong reactions in cases where the stakes are small. And committed act consequentialists will likely not share my reaction to this case or others like it; one philosopher's modus ponens can be another's modus tollens. In the end, this general critical strategy—that is, appealing to counter-intuitive cases as alleged counterexamples to consequentialist theories—will inevitably remain unconvincing to those with strong pre-existing theoretical commitments.

Singer frames a similar observation in this way:

> Most of the criticism [of act utilitarianism] has been inconclusive because it has consisted of the outlining of unusual situations, in which the application of act-utilitarianism is said to give results which conflict with our "ordinary moral convictions." This method of argument can never move anyone who has greater confidence in the act-utilitarian principle than in his "ordinary moral convictions." Whenever the conflict is a real one, and not merely an apparent conflict, dependent on the omission of factors which the act-utilitarian can and should take into account, the genuine act-utilitarian will be prepared to jettison his "ordinary moral convictions" rather than the principle of act-utilitarianism. (1972, 94)

Singer suggests that appealing to intuitions about cases in which a consequentialist account of promising leads to counter-intuitive results will get us nowhere, because the utilitarian will never accept these intuitions. We must also note that the critic of consequentialism will not be prepared to jettison her intuitions about cases in order to accommodate a theoretical principle that she does not antecedently find compelling—for it is (p. 296) likely that the act utilitarian's theoretical commitments are at least partially

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

grounded in intuitions that are not shared by nonutilitarians (e.g., about whether there is a moral difference between doing vs. allowing harm, about the sources of intrinsic value, etc.).

3.1.3. Underlying Theoretical Disputes

Consequentialists and their critics seem to be at an impasse: nonconsequentialists offer alleged counterexamples that they find compelling, and consequentialists insist that any such cases are unproblematic (either because promise-keeping is obligatory on consequentialist grounds in that case or because the intuition that promise-keeping is obligatory in that case is misguided). One way to move forward is by defending a general philosophical account of which kinds of intuitions are a reliable starting point for moral theorizing, and then arguing that the intuitions elicited by nonconsequentialist counterexample cases (such as my airplane case) are or are not among the class of reliable ones (e.g., because they are or are not widely shared, or are self-evident, etc.). However, I think that a more productive approach to moving past this impasse requires better understanding the underlying and often unarticulated theoretical disputes between consequentialists and nonconsequentialists about the nature and stringency of promissory obligation.

First, consequentialists and nonconsequentialists fundamentally disagree about *what kind of reason* grounds the obligation to keep a promise. The consequentialist presumes more generally that all moral reasons are grounded in the beneficial consequences of actions, while critics of consequentialism disagree and suggest that distinctively promissory obligations must be grounded in something that is less contingent and more essential to the making of the promise (such as an invitation to trust in the promisor, the transfer of a normative power to exercise decision-making authority over the promised action, contractualist agreements that no reasonable person could dissent from, etc.).[12] This is related to the more general dispute between consequentialists and nonconsequentialists over rights. Nonconsequentialist theories of promissory obligation often presume that A's promising B to Φ at time *t* involves A granting or transferring to B the right to decide whether A will Φ at *t*. Under this picture, wrongfully breaking a promise involves a rights violation. On straightforward consequentialist accounts—which do not take rights seriously—breaking a promise must be wrong for a different sort of reason, a reason that does not have to do with anything that A owes to B specifically.

Second, nonconsequentialists presume that A's promising B to Φ and B's accepting this promise generates a pro tanto moral obligation for A to Φ. Pro tanto obligations are *strict*, or such that you are required to act on them when all else is equal, and you go wrong in some way if you do not. If you have promised me that you will meet me for lunch at Spiral Diner at noon, you *must* do so, unless you have a good excuse. Contrast this with mere reasons, which are not strict; you have a reason to eat lunch (p. 297) at Spiral Diner because the food is delicious, but do not go wrong if you do not act on this. Pro tanto obligations are also *overridable*: sometimes all else is not equal, and the demand is outweighed or undermined by other demands. All else is not equal when excusing circumstances

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

arise, such as needing to break a promise to satisfy a more important conflicting obligation or discovering that the promise was based on deception.[13]

On the surface, counterexamples to consequentialist accounts of promising are a species of a much broader objection to consequentialism: namely, that consequentialism has unintuitive implications and requires us to act against the demands of common-sense morality. However, these objections also reflect deeper theoretical disputes about whether promises generate strict pro tanto moral obligations, about whether promise-breaking involves rights violations, and about whether the moral force of promising stems exclusively from the contingent benefits of promise-keeping or (also) depends on some other non-contingent source. Adjudicating this dispute requires more than intuitive reflection about whether promise-keeping is obligatory in particular concrete or abstract cases. Rather, it requires a broader theoretical assessment of the plausibility and explanatory power of strict pro tanto moral obligation as a conceptual category, and of what the source of such obligations might be. Reflection on promissory obligation will surely play an important role in this discussion. But it cannot be the only focus.

3.2. Are Act Consequentialist Theories Circular?

Another strategy for criticizing act consequentialist accounts of promising involves claiming not that they lead to counterintuitive results, but that they are inherently circular (or even self-refuting).[14] Because this critical strategy does not appeal to intuitions about cases, it avoids the impasse encountered earlier. The basic worry is that there is a good utilitarian reason to make and keep a promise to Φ (as opposed to simply stating an intention to Φ and then so acting) only if this leads the promisee to form a utility-increasing expectation that the promisor will Φ. But the promisee will form such an expectation only if she antecedently believes that the promise obligates the promisor to Φ, thereby making it likely that the promisor will Φ. And the promisor will be so obligated (and will accordingly be likely to Φ) only if the promisee expects the promisor to Φ. But the promisee will not have this expectation unless she believes that promise-keeping (p. 298) is obligatory for the promisor. And promise-keeping will not be obligatory unless the promisee forms an expectation: the theory runs into a circle.[15]

I do not think that circularity as such is a serious problem for the act consequentialist. For act consequentialists need not be committed to the claim that expectations *alone* explain why we ought to keep promises. Rather, they can bolster their accounts by appealing to utility-maximizing social conventions. If creating and generally adhering to a social convention of promising that enables us to form promissory expectations makes us all better off, this convention will itself be justified on utilitarian grounds. The convention gives promisees reason to expect that promisors will perform as promised, thus getting us out of the circle.

However, this solution suggests a new worry for the act consequentialist, which is closely related to the previous objection. While adhering to a promising convention that demands satisfying promisee expectations will *generally* lead to the best overall consequences, this

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

won’t always be so. For there will surely be some cases—and perhaps many—in which violating the norms of the convention and leaving promissory expectations unfulfilled leads to better results than does keeping the promise. Promisors will have to assess whether the generally valuable promising convention ought to be followed in any particular case. While the convention enables expectations to be formed and thus makes promising possible, it does not provide an act consequentialist justification for promise-keeping in particular cases; only direct appeal to the good or bad consequences of promise-keeping can do that. And direct assessment by a promisor of whether keeping a particular promise leads to the best results overall is both burdensome and potentially unreliable, biased as we all tend to be in favor of our own interests. This concern motivates some to move to rule consequentialism instead, discussed in section 5.

3.3. Do Act Consequentialist Accounts Prove Too Much?

Finally, we should consider an interesting and underappreciated worry raised by Michael Cholbi (2014). Cholbi argues that the most plausible act utilitarian accounts ground promissory obligation in the value of assuring the promisee that an action that the promisee takes to be valuable on independent grounds will be performed. In most cases, keeping such promises maximizes utility and is accordingly obligatory. But in many cases in which keeping a particular promise maximizes utility, *making* that promise—and thereby creating an expectation which it is then utility-maximizing to fulfill—will *itself* be utility-maximizing compared to refraining from making the promise. Cholbi offers an example of Matilda and Ned, a couple who are perfect for each other but are hesitant to make a permanent commitment due to their past romantic histories. They would both be much happier if they married each other, as their relationship would feel more secure. Marriage involves making promises to each other; it follows that they (p. 299) are obligated on utilitarian grounds to make marriage promises to each other. Cholbi concludes that it follows from this that “a wide range of seemingly discretionary human choices, choices usually seen as matters of personal prudence rather than impersonal morality, become matters of moral obligation on a utilitarian view” (264).[16]

But this is not how we generally construe promise-making: philosophers and lay people alike widely presume that promises are optional or discretionary, and that we are generally not morally required to make them in the first place, even if we are required to keep them once made.[17] Utilitarianism is often accused of being overly demanding,[18] and Cholbi argues that this is an additional way in which utilitarianism demands too much: if the value of assurance grounds a utilitarian obligation to *keep* promises, it will frequently also ground a utilitarian obligation to *make* promises. The utilitarian defense of promise-keeping comes saddled with a commitment to the obligatoriness of promise-making.

I suspect that the situation is not as dire as Cholbi makes it out to be, because I doubt that promise-making will often successfully increase utility if the promisor is not already voluntarily inclined to make the promise. Cholbi presumes that the utilitarian value of promising stems from the promisor’s satisfying the desires of the promisee. We often desire that others act in certain ways toward us because they have voluntarily chosen to,

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

and not because they are obligated to. For example, I might prefer that a dinner guest say nothing rather than compliment me on the meal I have served simply because she believes that this is what etiquette demands. Similarly, Ned might prefer that Matilda refrain from marrying him rather than marry him because she believes that doing so is morally obligatory. In such a case, Matilda's utilitarian obligation to make a marriage promise is narrower than Cholbi makes it out to be: she is obligated to promise *for discretionary reasons.* Either she is already motivated by discretionary reasons, in which case the obligation is inert and can do no motivational work, or she is not already motivated by discretionary reasons, in which case making the promise would not maximize utility after all.

However, while I am not convinced that act utilitarians will be faced as many morally obligatory promises as Cholbi suggests, I do not doubt that there will be some cases in which promise-making is morally obligatory on act utilitarian grounds (because the promisee does not have the preference just described, or because the benefits of making the promise outweigh the promisee's preference that the promise be made only if it is voluntary). Act consequentialists must accept that this is a feature of their view that **(p. 300)** many will find counterintuitive.[19] This is the flip-side of the worry from section 3.1 about whether act consequentialist views require promise-keeping in too few cases: perhaps these views also require promise-making in too many cases. And a similar theoretical impasse over conflicting intuitions is likely to arise.

Ultimately, the three objections I have discussed in this section show us that act consequentialist theories of promissory obligation are always to some extent *revisionary* of our common presumptions about promising*:* they require jettisoning the presumption that promises create strict pro tanto obligations, admitting that we are at least sometimes not morally obligated to keep our promises, and accepting that promise-making is sometimes morally required. These revisions do not by themselves conclusively establish that simple act consequentialist views are misguided; maybe our common presumptions about promising are problematic instead. But the need for revisions is a cost for simple act consequentialist theories. In light of this, it is worth looking at more complex act consequentialist accounts to see if they can avoid such revisionism.

4. Alternate Act Consequentialist Accounts

4.1. Multiple Intrinsic Goods

So far, I have been assessing act consequentialist accounts that aim to maximize a single good (such as pleasure, or well-being construed more broadly). Can consequentialist views that allow for multiple intrinsic goods offer a more plausible or less revisionary explanation of promissory obligation? Philip Pettit (2018) argues that we should expand our conception of what can properly count as the consequences of an action for the purposes of weighing outcomes in a consequentialist theory. He includes goods that are constitutively part of the action rather than a causal effect of it, and he argues that the agent's

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

motivation or disposition in acting can partially determine the value of the action. In making this argument, Pettit presumes that a plausible consequentialist theory should include multiple intrinsic goods, and he suggests in passing that promise-keeping could be one of them. A consequentialist who accepted that promise-keeping was an intrinsic good could avoid many of the counterexample cases pointed to in section 3.1. For the intrinsic value of keeping of the promise itself would often tip the scales in favor of promise-keeping even in cases in which it appeared at first glance that more good could be done by promise-breaking (although there would still be some cases in which breaking the promise was necessary for the maximization of other, weightier intrinsic goods).

(p. 301) However, Peter Railton (2018) offers a reply in which he explains why consequentialists should be hesitant to expand their list of fundamental intrinsic goods to include things like promise-keeping. Railton presumes that a fundamental feature of consequentialist theories is an explanatory structure that "account[s] for the various forms of moral appraisal, such as moral rightness and moral goodness or virtue, in terms of tendencies to realise fundamental, intrinsic non-moral good" (34).[20] He notes that pleasure is a fundamentally nonmoral good, as is friendship. But he argues that the value of keeping a promise is fundamentally moral; promise-keeping is good only because and to the extent that it manifests moral goods like integrity, loyalty, or gratitude. If Railton is right about this, then accepting promise-keeping as an additional intrinsic good imposes a serious theoretical cost on a consequentialist theory: it changes the basic explanatory structure of the theory.

Advocates of promise-keeping as an intrinsic good could respond in (at least) two ways. First, they could claim that something other than the reduction of the moral to the nonmoral is fundamental to the structure of consequentialist theories. For example, Pettit (1997) has argued that consequentialism essentially requires that agents *promote* values (rather than *honoring* them).[21] If this is the right way to understand what is distinctive about consequentialism, this lessens the theoretical cost of accepting multiple fundamentally moral intrinsic goods. Second, they could accept Railton's assumption about the explanatory structure of consequentialism and seek to identify other, uncontroversially nonmoral goods that the value of promise-keeping bottoms out in, such as the value of building trust between strangers, or of sustaining a personal relationship, or of enabling group cooperation.[22] Ultimately, the core issues raised by this debate—that is, whether promise-keeping can plausibly be considered an intrinsic good, and whether the value of promise-keeping is fundamentally moral or nonmoral—are underexplored and well worth further consideration.

4.2. Directedness and Agent-Relative Values

Agent-neutral consequentialist accounts struggle to explain how promisees should be prioritized over others in the provision of benefits because of their status as promisees. As Atiyah puts it, "the felicific calculus simply fails to pay due regard to the identity of the parties concerned" (1981, 70). Ross takes the failure to accommodate the personal character of duty to be a core defect of utilitarian views: while the utilitarian presumes

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(p. 302) that "the only morally significant relation in which my neighbors stand to me is that of being possible beneficiaries of my action" (1930, 19), we must also recognize the moral significance of other relationships such as promisor and promisee, and we should not be neutral between benefitting a promisee or benefitting a third party. Could an agent-relative version of act consequentialism accommodate this insight?

Michael Smith takes this worry about agent-neutrality very seriously, noting that consequentialist accounts of promissory obligation grounded in agent-neutral values entail that "an agent's obligations turn out not to be especially targeted on the promises that *he* makes. This is remarkable because promissory obligation, at least as ordinarily understood, *is* so targeted" (2011, 209, emphasis in original). This version of the directedness objection takes this concern a step further: not only does agent-neutral consequentialism imply that promisors should not automatically prioritize their promisees over third parties, but it also implies that promisors should not automatically prioritize keeping the promises that they themselves have made.

In response, Smith suggests turning to an agent-relative consequentialist theory.[23] A common argument against agent-relative theories (such as hedonistic egoism) is that they are incoherent: if you have granted that pleasure is intrinsically valuable, it is inconsistent to grant that *my* pleasure is valuable and helps determine what I ought to do without also granting that *your* pleasure is valuable and also helps determine what I ought to do. Smith argues that this objection equivocates on the meaning of "valuable," which is fundamentally relational: my pleasure is *valuable-to-me*, while your pleasure is *valuable-to-you*. It follows that "your pleasure simply isn't relevant to my obligations, nor is mine relevant to yours, because though reason is on the side of my desiring or aiming at my pleasure, it isn't on the side of my desiring or aiming at yours, and vice versa" (212).

Having made room for an agent-relative theory, Smith considers how "a consequentialist who has an open mind about the possibility of there being relative values that explain the distinctive [i.e., directed or relational] nature of promissory obligation" could construct a theory (215). He starts with Scanlon's nonconsequentialist expectationalist theory, which draws on contractualism to ground promissory obligation in the value of satisfying expectations that you have deliberately raised in others. Smith concludes that Scanlon's view is best understood as implicitly consequentialist, "because the facts about obligation that it describes are all ultimately grounded in, and thus reduce to, facts about the relative value of agents meeting the reasonable expectations that they knowingly create" (214).

This approach offers a theory that is technically consequentialist and can avoid revisionary conclusions about the directedness of promissory obligation. But Smith's agent-relative consequentialist defense of promising is very revisionary in other ways. Many philosophers presume that agent neutrality is a central and distinctive feature of (p. 303) consequentialism.[24] Switching to an agent-relative account requires revising a common assumption about the structure of consequentialism, which will likely be taken by those who accept this assumption to be a serious theoretical cost. Other sophisticated versions of act consequentialism that avoid counterintuitive intuitions about the nature and strin-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

gency of promissory obligation are likely to have to make a similar trade-off: additional theoretical complexity (and arguably, implausibility, or a lack of distinctiveness compared to deontological theories) is the price to pay for adherence to common-sense moral assumptions. As Brad Hooker notes, theories that attempt to force act consequentialism to have the same results as deontological theories must generally "*postulate* so much (e.g., agent-relative value, desert, etc.) in order to squeeze everything into act-consequentialism that the resulting theory is left *explaining* very little" (2011, 214).

5. Rule Consequentialist Accounts of Promising

The challenges act consequentialist theories face in explaining promissory obligation have led some to embrace rule consequentialism instead. In "Two Concepts of Rules," John Rawls articulates a basic rule utilitarian theory of promissory obligation.[25] Rawls notes that keeping promises will not always maximize utility. But he argues that it does not follow from this that utilitarians should refrain from keeping such promises, because the utilitarian justification does not apply to individual promises themselves. Rather, our *practice* of promising is justified on utilitarian grounds, and the practice has its own internal rules (analogous to the rules of a game like baseball or chess) that demand that promises are generally kept. We can deliberate about how to construct this practice (e.g., what conditions should count as excuses, how we should penalize promise breakers, etc.). (p. 304) But once we are participating in the practice, we cannot engage in such deliberation, because "it is a mistake to think that if the practice is justified on utilitarian grounds then the promisor must have complete liberty to use utilitarian arguments to decide whether or not to keep his promise. The practice forbids this general defense; and it is a purpose of the practice to do this" (1955, 16).

Other rule consequentialists have refined this approach.[26] I will focus on Brad Hooker's (2011) version of rule consequentialism, since he has offered the most explicit, sophisticated, and detailed defense of promissory obligation in particular on rule consequentialist grounds.[27] According to Hooker's version of rule consequentialism:

> An act is wrong if it is forbidden by the code of rules whose internalization by the overwhelming majority of everyone everywhere in each new generation has maximum expected value in terms of well-being, with some priority for the worst off. (222)

The code of rules is fixed not by the rules people actually have internalized, but by the rules that would lead to the best consequences if people were to internalize them. Hooker proposes that such a code would include a rule that generally forbids us from breaking promises. Moreover, he claims that the practice of promising demanded by this rule would meet six theoretical desiderata that are typically thought the province of deontological theories of promising.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The first of these desiderata is most relevant for our present discussion. It states that promissory obligation is not contingent on benefits to the promisee or others.[28] This is in part because Hooker shares the intuition that promises create strict pro tanto obligations; he presumes that "the idea that promising creates duties to do things other than maximize good … and the idea that promising involves transferring a right from the promisor to the promisee are ideas that I think any moral theory must accommodate if it is to be plausible" (240). Hooker argues that a rule according to which promises are binding only when keeping them is beneficial would be nonideal, because it would induce insecurity in the promisee. Because "keeping a promise is very often disadvantageous to the promisor," promisors are likely to weasel out of promises whenever they can; it follows that "promisees need assurance that promisors are not left with much room for talking themselves into believing that their promises aren't binding" (250).

Hooker claims that if the promising rule permitted us to break promises whenever doing so failed to be beneficial, promisees would not have such assurance. His reasoning is not explicit here, but he seems to assume that if the promising rule contained a loophole allowing promises to be broken whenever they were not beneficial, self-interested promisors would too often exploit this loophole by convincing themselves that their (p. 305) promise lacked benefit and was therefore permissible to break, even when this was not the case. And the existence of such an easily exploitable loophole would so decrease promisees' confidence that their promises would be kept that our valuable practice of promising would be undermined. It follows that the code of rules whose internalization would best maximize expected well-being would not permit promise-breaking whenever doing so led to increases in benefit.

6. Criticisms of Rule Consequentialist Accounts

6.1. Avoiding Collapse

Rule consequentialism is often accused of being subject to worries about incoherence or collapse.[29] If rule consequentialism is overridingly committed to maximizing expected value, but insists that we should stick to some rule even when doing so does not maximize expected value, the position seems incoherent. But if rule consequentialism maintains coherence by allowing that we should maximize expected value in any case in which doing so is against one of the established rules, then the rule will need to include so many exceptions that the view is extensionally equivalent to act consequentialism.

Hooker's version of rule consequentialism avoids the collapse worry because of his constraint that the moral rules be internalized. The costs associated with internalization of the rules are factored into the calculation of expected value. As rules become more complex or demanding, or as they multiply in number, the cost of internalizing them increases. There will come a point at which the additional cost of internalizing extremely complex/demanding/varied rules outweighs any increases in expected value afforded by such

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

rules. So it is extremely unlikely that the rules will become so complex that they are extensionally equivalent to act consequentialism.

6.2. A Challenge for Hooker's Account

Hooker's argument presupposes that promisee assurance is so important for the success of the overall promising practice that it is worth protecting even at the cost of requiring promise-keeping in particular cases that do not maximize well-being. It also presupposes that promisors are so motivated to get out of their burdensome promissory obligations that they will be highly unreliable at assessing when a promise fails to benefit the (p. 306) promisee; we cannot give them any leeway in so assessing, lest we open the floodgates to promisors wrongly assuming their promises are of little or no benefit.

However, I worry that these presuppositions prove too much. Even understood as strict pro tanto obligations, promises are overridable and may be permissibly broken in a wide variety of cases, such as when keeping the promise is incompatible with satisfying a more important obligation or was based on a false claim. If promisors are highly motivated to get out of their burdensome obligations, won't they *also* be overly quick to wrongly assume that a promise was premised on false information, or that keeping it conflicts with satisfying a more important obligation? And won't this too lead to promisee insecurity and lack of assurance? The same reasoning Hooker employs to justify constraining the promising rule so that promises must be kept even when the promise is of no benefit to the promisee might also justify constraining the rule so that promises must be kept even when they conflict with a more important obligation or are grounded in a false belief.

Hooker presumes that maintaining promisee security is so important that it is worth the cost to overall expected well-being of occasionally requiring actions that do not benefit anyone in particular cases. If promisee security really is so essential to the successful practice of promising, it is at least plausible that maintaining promisee security is *also* worth the cost to overall expected well-being of occasionally acting on a promise grounded in a false belief, or of failing to satisfy more important obligations in particular cases. But this is a bad result; we don't want our promising rule to be so constrained. It seems that Hooker's view proves too much: a strict promise-keeping rule that forbids promise-breaking in the cases Hooker wants to accommodate risks also forbidding promise-breaking in cases in which there should be an excuse in place. If this is correct, then Hooker's rule consequentialism has revisionary implications in a way similar to that of act consequentialist views.

Hooker could potentially respond to this objection in the following way. Perhaps third parties are better than promisors are at determining when it is best to break a promise, all-things-considered. An optimific rule about when to blame promise-breakers would track with these judgments, and the threat of potential sanction would in many cases prevent promisors from too readily breaking promises in their own favor. Internalizing a permissive promising rule that allows for breach of promise in a fairly wide range of cases—*paired with* a rule requiring third-party sanction if the promisor misjudged—might be

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

more optimific than internalizing a restrictive promising rule that forbids promise-breaking in most cases. More reflection on whether rule consequentialism can accommodate a rule requiring promise-keeping in all and only those cases demanded by common-sense morality is needed.

7. Upshots and Future Directions

Answering the question of whether consequentialism can accommodate promissory obligation requires more than just reflecting on our immediate intuitions about whether (p. 307) promise-keeping is morally required in some hypothetical, non-utility-maximizing case. We must also consider a range of questions about the nature of promising as a moral phenomenon, including whether promises yield strict pro tanto obligations (and how these obligations should be understood); what the excusing conditions on promises are (and whether they line up with the cases in which consequentialist theories predict promise-keeping is not required); how and whether the strength of promissory obligation varies (and whether consequentialist theories give us a good explanation of this—see note 4); whether promise-making is ever morally obligatory (and if consequentialism problematically implies the wrong answer to this question); and so on. The answers to these questions can help shape our desiderata for any plausible theory of promissory obligation, be it consequentialist, deontological, or other.

Careful reflection on the relationship between consequentialism and promises also sheds light on the role that appeal to intuitions regarding (or speculation about the results of) imagined cases can and should play in the assessment of a moral theory more generally. Specifically, it illustrates the need for cases with a high degree of specificity whenever possible, especially for topics (such as assessing the results of promise-keeping or -breaking) that involve a large degree of uncertainty. It also illustrates the limits of this sort of reflection in isolation: intuitions about cases are likely to clash on the basis of existing theoretical and intuitive commitments, and they must be bolstered by appeals to independent criteria like those addressed in the previous paragraph.

Ultimately, I am skeptical that a traditional act or rule consequentialist theory will be able to accommodate all of our theoretical desiderata and pretheoretical intuitions about the robustness and breadth of promissory obligation. Most likely, there will always be some case in which the consequentialist must admit that a promise ought not be kept, despite deontologists' intuitions to the contrary. Consequentialists have a variety of resources to limit such cases, or minimize their forcefulness. But if counter-intuitive cases remain, this is a theoretical cost that consequentialism must deal with.

In light of this, consequentialists should continue to try to accommodate some version of promissory obligation to whatever extent they can, either by explaining how their views can accommodate a wider range of promissory obligation than it may first appear, or arguing that the cases that the views cannot explain are not problematic. One way forward is looking beyond straightforward utilitarian theories and seeing whether more sophisticated versions of consequentialism that are independently plausible can better accommo-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

date promissory obligation. Such accommodation would be a point in favor of a theory, and philosophers developing such theories would do well to explicitly discuss what their theories imply about promise-keeping.

That being said, I do not think that an inability to accommodate all of our intuitions about promise-keeping is a decisive strike against consequentialism as a moral theory. For our intuitions about the strength and breadth of promissory obligation are not independently robust enough to by themselves merit complete rejection of an otherwise powerful moral theory. Rejections of utilitarianism as a moral theory on this basis move (p. 308) too quickly: in the end, whether consequentialism stands or falls will have to be decided on broader grounds than this.[30]

References

Árdal, Páll S. 1976. "Promises and Reliance." *Dialogue* 15, no. 1: 54–61.

Atiyah, P. S. 1981. *Promises, Morals and the Law.* Oxford: Oxford University Press.

Bentham, Jeremy. (1840). 1977. *A Comment on the Commentaries and A Fragment on Government*. Edited by J. H. Burns and H. L. A. Hart. London: Athlone Press.

Brandt, Richard. 1959. *Ethical Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Brandt, Richard. 1992. *Morality, Utilitarianism, and Rights*. New York: Cambridge University Press, 1992.

Carson, Thomas L. 2005. "Ross and Utilitarianism on Promise Keeping and Lying: Self-Evidence and the Data of Ethics." *Philosophical Issues* 15, no. 1: 140–157.

Cholbi, Michael. 2002. "A Contractualist Account of Promising." *Southern Journal of Philosophy* 40: 475–491.

Cholbi, Michael. 2014. "A Plethora of Promises—Or None at All." *American Philosophical Quarterly* 51, no. 3: 261–272.

Darwall, Stephen. 2003. "Theories of Ethics." In *A Companion to Applied Ethics*, edited by R. G. Frey and Christopher Heath Wellman, 17–37. Malden, MA: Blackwell.

Friedrich, Daniel, and Southwood, Nicholas. 2009. "Promises beyond Assurance." *Philosophical Studies* 144, no. 2: 261–280.

Gibbard, Allan F. 1965. "Rule-Utilitarianism: Merely an Illusory Alternative?" *Australasian Journal of Philosophy* 43, no. 2: 211–220.

Gill, M. B. 2012. "The Non-Consequentialist Moral Force of Promises: A Response to Sinnott-Armstrong." *Analysis* 72, no 3: 506–513.

Goldman, Holly S. 1974. "David Lyons on Utilitarian Generalization." *Philosophical Studies* 26, no. 2: 77–95.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Hooker, Brad. 2011. “Promises and Rule-Consequentialism.” In *Promises and Agreements*, edited by Hanoch Sheinman, 237–254. Oxford: Oxford University Press.

Howard-Snyder, Frances. 1994. “The Heart of Consequentialism.” *Philosophical Studies* 76, no. 1: 107–129.

Hume, David. (1741). 1978. *A Treatise of Human Nature*. 2nd ed. Edited by L. A. Selby-Bigge and P. H. Nidditch. Oxford: Clarendon Press.

Kolodny, Niko, and Wallace, Jay. 2003. “Promises and Practices Revisited.” *Philosophy & Public Affairs* 31, no. 2: 119–154.

Liberman, Alida. 2015. *The Mental States First Theory of Promising.* PhD diss., University of Southern California.

Lyons, David. 1965. *Forms and Limits of Utilitarianism*. Oxford: Clarendon Press.

McNaughton, David, and Rawling, Piers. 1992. “Honoring and Promoting Values.” *Ethics* 102, no. 4: 835–843.

Mill, John Stuart. (1863). 2001. *Utilitarianism*. 2nd ed. edited by George Sher. Indianapolis: Hackett.

(p. 309) Mukerji, Nikil. 2014. “Consequentialism, Deontology and the Morality of Promising.” In *Business Ethics and Risk Management*, edited by Johanna Jauernig and Christoph Lütge, 111–126. Dordrecht: Springer.

Narveson, Jan. 1971. “Promising, Expecting and Utility.” *Canadian Journal of Philosophy* 1, no. 2: 207–233.

Norcross, Alastair. 2011. “Act-Utilitarianism and Promissory Obligation.” In *Promises and Agreements*, edited by Hanoch Sheinman, 217–236. Oxford: Oxford University Press.

Owens, David. 2012. *Shaping the Normative Landscape.* Oxford: Oxford University Press.

Pettit, Philip. 1997. “The Consequentialist Perspective.” In *Three Methods of Ethics: A Debate,* edited by Marcia Baron, Philip Pettit, and Michael Slote, 92–174. Malden, MA: Blackwell.

Pettit, Philip. 2018. “Three Mistakes about Doing Good (and Bad).” *Journal of Applied Philosophy* 35, no 1: 1–25.

Pickard-Cambridge, W. A. 1932. “Two Problems about Duty (II).” *Mind* 41: 145–172.

Portmore, Douglas W. 2001. “Can an Act-Consequentialist Theory Be Agent Relative?” *American Philosophical Quarterly* 38, no. 4: 363–377.

Prichard, H. A. (1940) 2002. “The Obligation to Keep a Promise.” In *Moral Writings*, edited by Jim MacAdam, 257–266. Oxford: Clarendon Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Railton, Peter. 2018. “Toward a More Adequate Consequentialism.” *Journal of Applied Philosophy* 35, no 1: 33–40.

Rawls, John. 1971. *A Theory of Justice.* Cambridge, MA: Harvard University Press.

Rawls, John. 2005. “Two Concepts of Rules.” *Philosophical Review* 64, no. 1: 3–32.

Robins, Michael H. 1976. “The Primacy of Promising.” *Mind* 85, no. 339: 321–340.

Ross, W. D. 1930. *The Right and the Good*. Oxford: Clarendon Press.

Ross, W. D. 1939. *Foundations of Ethics.* Oxford: Clarendon Press.

Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Shiffrin, Seana. 2008. “Promising, Intimate Relationships, and Conventionalism.” *Philosophical Review* 117, no. 4: 481–524.

Shiffrin, Seana. 2011. “Immoral, Conflicting and Redundant Promises.” In *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon*, edited by R. Jay Wallace, Rahul Kumar, and Samuel Freeman, 155–178. Oxford: Oxford University Press.

Sidgwick, Henry. 1874. *The Method of Ethics*. Chicago: Chicago University Press, 1962.

Singer, Peter. 1972. “Is Act-Utilitarianism Self-Defeating?” *Philosophical Review* 81: 94–104.

Sinnott-Armstrong, Walter. 2009. “How Strong Is This Obligation? An Argument for Consequentialism from Concomitant Variation.” *Analysis* 69, no. 3: 438–442.

Smart, J. J. C. 1973. “An Outline of a System of Utilitarian Ethics.” In *Utilitarianism: For and Against*, edited by J. J. C. Smart and Bernard Williams, 3–76. Cambridge: Cambridge University Press.

Smith, Michael. “The Value of Making and Keeping Promises.” In *Promises and Agreements*, edited by Hanoch Sheinman, 198–216. Oxford: Oxford University Press, 2011.

Urmson, J. O. 1953. “The Interpretation of the Moral Philosophy of J. S. Mill.” *Philosophical Quarterly* 10: 33–39.

Zong, Desheng. 2000. “Agent Neutrality Is the Exclusive Feature of Consequentialism.” *Southern Journal of Philosophy* 38, no. 4: 676–693.

Notes:

([1]) There are also straightforward consequentialist assessments of whether rules or practices lead to the best overall consequences, including rules about promise-keeping. (This differs from rule consequentialism, which directly assesses rules in terms of their conse-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

quences and then determines the morality of actions by assessing whether they adhere to utility-maximizing rules.) For ease of discussion, I will focus on straightforward consequentialist assessment of acts.

([2]) For example, see Bentham (1840, 444–445), Mill (1863, chap. 5), Sidgwick (1874, book 3, chap. 6), Pickard-Cambridge (1932), Narveson (1971), and Singer (1972).

([3]) Hume argues that our natural selfishness and limited inclinations toward gratitude and generosity entail that we will be inclined to refrain from holding up our end of the bargain once we have received the benefit. Predicting this, others will not be likely to make bargains with us, and everyone will be worse off than they would be had we cooperated. We can avoid this problem only by creating and maintaining a social practice of promising. Because promise-breakers subject themselves "never being trusted again in case of failure" (*Treatise* 3.2.5.10, SBN 522), enlightened self-interest motivates people to keep their promises to ensure ongoing mutually beneficial exchange. Unlike most utilitarian accounts of promising, Hume's view is generally classified as conventionalist rather than expectationalist, which means that he takes the source of promissory obligation to be not the utility of satisfying the expectation, but rather the widely accepted and mutually beneficial convention itself.

([4]) Walter Sinnott-Armstrong argues for consequentialism by highlighting how consequentialist theories can easily explain the variable weight of promises. He offers cases in which the moral weight of the obligation to keep a promise corresponds to how harmful breach of promise would be, and he argues that "this correlation supports the hypothesis that the harms of violating it are what makes the moral obligation as strong as it is" (2009, 440). Michael Gill offers an alternate explanation, arguing that the strength of a promissory obligation varies not according to the degree of harm imposed, but according to how much the promisee values the promise being kept (2012). Nikil Mukerji offers a different reply, claiming that the strength of promissory obligation depends on the amount of harm that breach would cause the promisee in particular, rather than on the amount of general harm that would result (2014).

([5]) Ross's discussion of promise-breaking in *The Right and the Good* is sometimes taken to be a direct argument against utilitarianism, but it is better understood as an argument against the self-evidence of the utilitarian position. See also Ross's defense of his intuitionist account of promise-keeping against utilitarian critiques raised W. A. Pickard-Cambridge (Ross 1939, chap. 5). Thanks to Luke Robinson for discussion of these points.

([6]) In a paper critical of expectationalist views of promissory obligation, Páll Árdal (1976) suggests that deathbed promises are so unusual that they cannot tell us anything significant about our normal promising practices (although he goes on to note that there are utilitarian reasons to keep such promises, since a person's dying moments will be happier if she knows that deathbed promises are generally kept). However, this is not a compelling defense, since it will not cover all cases. And as Atiyah (1981) notes, secret deathbed promises remain problematic for consequentialists.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([7]) Ross uses “prima facie” in roughly the way that contemporary philosophers use “pro tanto.”

([8]) Norcross argues that this need not involve problematic self-deception, but grants that there remains “tension between the belief that x is morally required and a commitment to doing something other than x (such as keeping a promise)” (2011, 235). He suggests that we can ease this tension by rejecting the idea that consequentialism demands that we always maximize the good, and instead accepting scalar consequentialism, which he defends elsewhere. According to this view, in keeping the promise to his grandfather John does not do the morally *best* thing, but he has not failed to fulfill a moral requirement.

([9]) Thomas Carson articulates this worry as follows: “Ross’s second example presupposes that our judgments about the values of the consequences of actions have a far greater degree of precision than they, in fact, have. I cannot think of a case in which one course of action will produce exactly 1000 units of good and an alternative course of action which involves breaking a promise will produce exactly 1001 units of good. We need concrete examples against which to test our intuitions” (2005, 145).

([10]) This is especially so given the possibility of unexpected “chain reactions” (e.g., breaking a promise to re-pay a loan and then having bad credit for years); although “the risk [of a chain reaction] is no doubt a very small one in most circumstances … the consequences of such a breakdown would be so very great that it is undesirable to take such risks unless the gain manifestly outweighs the dangers” (Atiyah 1981, 79).

([11]) T. M. Scanlon (1998) describes a promise that does not raise expectations (your “Profligate Pal” asks you for a loan, and you give him the money without any expectation of repayment), but we have the intuition that keeping the promise is morally obligatory anyway, because it ought to be kept for independent moral reasons (i.e., the Pal should return the money because he feels gratitude toward you for a gift disguised as a loan). Assume that such alternate explanations do not apply in my case.

([12]) See Friedrich and Southwood (2009), Shiffrin (2008) and Owens (2012), and Cholbi (2002) for examples of views grounding promissory obligation in trust, the exercise of normative power, and contractualism, respectively.

([13]) See Liberman (2015, chap. 7) for a discussion of the excusing conditions on promises.

([14]) D. H. Hodgson articulates a strong version of the circularity objection in chapter 2 of *Consequences of Utilitarianism* (1967), in which he argues that utilitarianism is self-refuting, because a utility-maximizing practice of promising would be unable to get off the ground in a fully utilitarian world. While Hodgson’s argument was influential when it was first published, it has been subject to a number of forceful criticisms that have seriously blunted its force; for example, see Narveson (1971) and Singer (1972).

([15]) For examples of the circularity objection to expectationalist accounts of promising, see Prichard (1940) and Robins (1976). Kolodny and Wallace (2003) raise a circularity ob-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

jection to Scanlon's (1998) contractualist expectationalist account of promising and respond by developing a hybrid expectationalist/conventionalist view.

([16]) Cholbi also argues that utilitarianism implies that we will sometimes be obligated to promise to do what is antecedently obligatory for us, but that such redundant promises are unnecessary, and perhaps a sign of an agent's weakness. However, I am not convinced that redundant promises are problematic. See Shiffrin (2011) for a discussion of how redundant promises might make an imperfect duty more determinate, counterfactually commit a promisor to an action, or give greater weight to the promisee's perspective in deliberation as a form of partiality.

([17]) Cholbi does not suggest that common-sense morality entails that there is *never* a moral obligation to make promises. For example, some role relationships (such as friendship or a romantic partnership) might require that the parties at least occasionally make promises to each other to properly fulfil their roles.

([18]) See Sobel, Chapter 11, this volume.

([19]) Note that this same counterintuitive result might also apply to deontological theories (such as Rossian pluralism) that require maximization of benefit whenever possible as one of their pro tanto moral rules.

([20]) Stephen Darwall articulates a similar idea in this way: "Consequentialist moral theories start with a *non-moral value theory* …. Consequentialist moral theories all agree that the moral rightness and wrongness of acts are determined by the non-moral goodness of *relevant consequences*" (2003, 27).

([21]) To honor a value is to remain committed to realizing that value yourself (even if this leads to less of the value being instantiated in the world); to promote a value is to pursue its realization impartially in both yourself and others. Pettit argues that "the consequentialist says, first, that values determine rightness in the promotional, not the honoring way" (129).

([22]) Thanks to Josh Crabill for this suggestion.

([23]) See also Hurley, Chapter 2, this volume.

([24]) For example, McNaughton and Rawling argue that "the distinction between consequentialism and deontology is best drawn, we maintain, in terms of the distinction between the agent-relative and the agent-neutral" (1992, 836). Howard-Snyder (1992) argues that deontological theories can be agent-neutral, but insists that consequentialist theories cannot be agent-relative. Pettit argues that a key characteristic of consequentialism is that "the values which determine rightness are all neutral values, not values that have a distinctively relativized reference" (1997, 129). And Zong (2000) defends the claim that consequentialism (as a criterion of rightness, rather than as a decision procedure)

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

must be agent-neutral with respect to both agency and values. For an argument that act consequentialism is compatible with agent-relativity, see Portmore (2001).

([25]) Rawls does not use this terminology, which was not coined until Brandt (1959, chap. 15). In *A Theory of Justice* (1971), Rawls argues that fairness requires us to act in accordance with the just institutions and practices whose benefits we voluntarily accept. To fail to do so is to unfairly (and, ipso facto, morally problematically) free-ride on others' participation in the practice. Promising is a just and valuable practice, and making a promise involves voluntarily participating in that practice. It follows that promise-keeping is morally obligatory, lest one unfairly free-ride on a beneficial and voluntary social practice.

([26]) For example, see Urmson (1953) and Brandt (1959, 1992).

([27]) This discussion of Hooker's view and my criticisms of it (in section 6.2) are also addressed in Liberman (2015, chap. 2).

([28]) Hooker's other desiderata are that promissory obligations: "(2) are self-imposed (autonomy); (3) are backward-looking in the sense that they depend on events in the past; (4) are agent-relative; (5) confer rights on particular others; and (6) give only some others (the promisees) the status of being wronged if the promise isn't kept" (249).

([29]) See Lyons (1965, 182–195) for a classic articulation of this objection applied to promises in particular; see Gibbard (1965) and Goldman (1974) for general replies to Lyons's argument. See also the discussion in Hooker, Chapter 23, this volume.

([30]) Thanks to Eric Barnes, Philippe Chuard, Josh Crabill, Justin Fisher, Steve Hiltz, Robert Howell, Jennifer Matey, Luke Robinson, Steve Sverdlik, and Doug Portmore for helpful discussion of this paper.

Alida Liberman

Alida Liberman is Assistant Professor of Philosophy at Southern Methodist University. Her main research interests are in theoretical and applied ethics, and she is particularly interested in how our attitudes and commitments affect what it makes sense for us to do. She has published papers about promises, vows, endorsements, and resolutions, as well as about a variety of topics in bioethics, and is currently working on a project about the ways in which it is wrong to make it harder for others to fulfill their obligations.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Supererogation and Consequentialism

Alfred Archer

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.17

Abstract and Keywords

The thought that acts of supererogation exist presents a challenge to all normative ethical theories. This chapter will provide an overview of the consequentialist responses to this challenge. I will begin by explaining the problem that supererogation presents for consequentialism. I will then explore consequentialist attempts to deny the existence of acts of supererogation. Next, I will examine a range of act consequentialist attempts to accommodate supererogation, including satisficing consequentialism, dual-ranking act consequentialism, and an anti-rationalist form of consequentialism. Finally, I will explore how indirect consequentialists have responded to this problem. Throughout the chapter, I will argue that in responding to the challenge of supererogation, consequentialists must choose between a more theoretically satisfying version of consequentialism and a form of consequentialism that is better able to accommodate our everyday moral intuitions and concepts.

Keywords: anti-rationalism, consequentialism, common-sense consequentialism, dual-ranking consequentialism, indirect consequentialism, maximizing consequentialism, moral obligation, supererogation

1. Introduction

IT is a recognizable feature of common-sense morality that some actions are supererogatory or beyond the call of duty. J. O. Urmson (1958) is credited with opening the contemporary discussion of the concept of supererogation in moral philosophy.[1] Urmson argued that for a normative moral theory to be acceptable it must make room for supererogation. He gave the following example to support this claim:

> We may imagine a squad of soldiers to be practising the throwing of live hand grenades; a grenade slips from the hand of one of them and rolls on the ground near the squad; one of them sacrifices his life by throwing himself on the grenade and protecting his comrades with his own body. (1958, 63)

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The soldier has clearly acted in a way that is morally good. However, his act should not be considered morally required. This gives us reason to accept that there are some acts that are morally good but that are not morally required.

According to Urmson, the existence of acts of supererogation presents a challenge for normative ethical theories. In order to be plausible, these theories should be able to accommodate this feature of common-sense morality. However, Urmson claimed that there is no straightforward way for any of the major normative theories of his time (Kantianism, consequentialism, and Moorean intuitionism) to accommodate (p. 270) supererogation. Nevertheless, Urmson (1958, 72) did hold that of the three, consequentialism was the best placed to do so.

Rather than investigating Urmson's claim about the comparative advantages of consequentialism here, this chapter will examine the comparative merits of the different responses that consequentialists have made to this problem. I will begin, in section 1, by explaining the problem that supererogation raises for consequentialism. I will then, in section 2, explore consequentialist attempts to deny the existence of acts of supererogation. Next, I will examine a range of act consequentialist solutions to the problem. In section 3, I will explore the satisficing consequentialist response to the problem. Then, in section 4, I will investigate the response to the problem offered by dual-ranking consequentialism. In section 5, I will look at a solution to the problem that involves reinterpreting supererogation. Finally, in section 6, I will explore how indirect consequentialists have responded to this problem. I will argue that in responding to the challenge of supererogation, consequentialists must choose between a more theoretically satisfying version of consequentialism and a form of consequentialism that is better able to accommodate our everyday moral intuitions and concepts.

2. The Problem for Consequentialism

According to consequentialists, whether or not an act is right or wrong is determined solely by the comparative value of that action's consequences compared to the alternative acts available. This broad definition leaves a range of important issues open. Most importantly, we need an account of how we determine which consequences are better than others. The most important question for present purposes arises once a way of ranking of consequences has been determined. Assuming such a ranking, how do we decide which acts are morally permissible and which acts are morally wrong?

The most straightforward answer to this question is the following:

> *Maximizing Consequentialism*: It is permissible for an agent A to perform act φ at time *t* if and only if there is no other act that A could perform at t that would bring about better consequences than φ.

In other words, at any given time the only act that is permissible from the acts that are available is the one that would bring about the best consequences. Or in the case where

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

two or more acts are tied for first place, then all and only those actions ranked in first place are permissible.

However, while this account offers a straightforward account of the connection between the evaluation of consequences and moral permissibility, it faces clear problems in accommodating acts of supererogation. To show why, I will first clarify exactly how the term "supererogation" should be understood. First, supererogatory acts are permissible to perform or omit. Consider the soldier in Urmson's example. He is (p. 271) permitted to act in this heroic way, but it would also be permissible for him not to do so. We can formulate this point in the following way:

> *Morally Optional*: If an act φ is supererogatory, φ is morally permissible, but φ is not morally required.[2]

Maximizing consequentialism can accommodate this aspect of supererogation. If there are two or more acts that are tied for first place in the rankings for the best consequences, then it will be permissible to perform either act. Both acts then are, in a sense, morally optional.

However, being morally optional is not sufficient for an act to be supererogatory. Imagine that I am buying an ice cream in a café and have to choose between two flavors I like equally and my choice will not affect anyone else. It is morally optional for me to choose either flavor, but neither choice would be supererogatory. What is missing from this example that means that neither act is supererogatory is that neither is better than the other from the moral point of view. In contrast, when we consider Urmson's example, the soldier can choose between two permissible options (jumping on the grenade or standing still), but one of these options is morally better than the other. In order for an act to be supererogatory then, it must also meet the following condition:

> *Morally Better*: If an act φ is supererogatory, φ is morally better than at least one other morally permissible alternative.[3]

A supererogatory act, then, is one that it is permissible both to perform and to omit and the performance of which is better than some other permissible alternative.

Maximizing consequentialism cannot accommodate this aspect of supererogation. The reason for this is that an act is only permissible according to maximizing consequentialism if there are no other available acts that would have better consequences. Or to put it another way, an act is only permissible if there are no other acts available to the agent that it would be morally better to perform.[4]

We can now see the problem that consequentialism faces in accommodating supererogation. According to consequentialism, the moral status of an action is fully determined by its consequences. Once consequentialists have provided an account of how (p. 272) the consequences of the various acts we could perform are to be ranked, they must then provide an account of how this ranking translates to moral concepts such as obligation, permissibility, wrongness, and indeed supererogation. The most straightforward answer is

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

maximizing consequentialism, according to which the only permissible acts are those that bring about the best consequences. However, this view leaves no room for actions that are morally better than another permissible action. There is then, no way for this view to accommodate supererogation.

This leaves the consequentialist with two options. First, bite the bullet and accept that their view cannot accommodate supererogation. Those who take this path typically seek to argue that denying the existence of supererogation is not as problematic as it may initially appear. Alternatively, find an alternative account of the connection between the evaluative ranking of actions and moral concepts such as obligation, permissibility, wrongness, and supererogation. I will first consider arguments from those who take the first option before considering a range of different approaches to the second.

3. Rejecting Supererogation

The most straightforward response that consequentialists can make to this problem is to reject the claim that acts of supererogation exist. This approach is endorsed by consequentialists such as Fred Feldman (1986) and Shelly Kagan (1989), who argue that we ought to always do the best we can from the moral point of view. As discussed in the previous section, if we accept that we ought always to do what is morally best, then this leaves no room for acts that are beyond the call of duty.

The advantage of this response is that it does not require the consequentialists to compromise their position. The disadvantage is that this response offers an account of morality which many will view as over demanding and which leaves no room for acts of supererogation. Kagan and Feldman offer a number of interesting responses to the counterintuitive implications of their views, in particular to the claim that such views are too demanding (for a discussion of the demandingness of consequentialism see Sobel, Chapter 11, this volume). In my view, these arguments should be assessed alongside an assessment of the plausibility of the views as a whole; therefore, I will not explore them in detail here. It is worth noting, though, that Kagan (1984) provides a number of responses to arguments that Heyd (1982) offers in support of the claim that acts of supererogation exist. Consequentialists might also draw on the arguments that others have made against the existence of acts of supererogation in order to support this approach (see Archer 2018, 1–4, for an overview).

A specifically consequentialist rationale for rejecting the existence of acts of supererogation is offered by Alastair Norcross (1997), who claims that consequentialists cannot give a satisfactory account of moral terms such as rightness, wrongness, goodness, duty, and permission and so should remove them from their moral lexicon. Instead, Norcross (1997) argues that consequentialists should talk only of alternative states of affairs as (p. 273) being better or worse.[5] Norcross (2006, 44) claims that this scalar consequentialism has an advantage over nonscalar forms of consequentialism when it comes to the problem of supererogation. Admittedly, scalar consequentialism will have to eliminate the concept of supererogation along with the concept of duty. If there are no duties, then there can be

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

no acts that go beyond those duties. However, the scalar consequentialist can accommodate the intuitions that would push us to accept the existence of acts of supererogation by explaining these "in terms of actions that are considerably better than what would be expected of a reasonably decent person in the circumstances" (Norcross 2006, 44). A maximizing consequentialist would have to claim that people have a *duty* to do what is best and as a result that this is demanded of them. Scalar consequentialism, on the other hand, does not demand any one course of action and so can allow that the best available act is not demanded of us.

Like the views of Kagan and Feldman, Norcross's view proposes a restructuring of our moral lives. However, rather than claiming that moral obligations are more radically demanding than we think, Norcross claims that morality does not involve any obligations at all. While we can evaluate alternative acts as better or worse, a scalar consequentialist will never translate this into a deontic judgement of obligation, wrongness, or permissibility. It is worth noting just how radical this view is. These concepts play an important role in our ordinary moral discourse. We might find it hard to imagine what our moral practices would look like without these concepts. More importantly, we might think that such practices would be impoverished in comparison to moral practices that can make room for these concepts (McElwee 2010a). Moreover, we may find it strange that moral philosophers would end up with a more restricted moral vocabulary than common-sense morality (McElwee 2010a, 314).

Moreover, there seem to be specifically consequentialist reasons to worry about this approach. As Gerald Lang (2013) notes, the evaluative assessments that consequentialists typically endorse are plausibly seen as generating reasons for action, which in turn are plausibly see as generating obligations. Similarly, as Katarzyna de Lazari-Radek and Peter Singer (2014, 334) note, the question of what we ought to do was viewed as the fundamental ethical question in one of consequentialism's foundational works, Henry Sidgwick's *The Methods of Ethics* (1907). It is fair to say that none of these points provide decisive arguments against scalar consequentialism. Like Kagan and Feldman's views, the only way to fairly assess this view is to evaluate the view's radical implications in comparison with the plausibility of the arguments offered in favor of it.

In this section we have examined one consequentialist response to the problem of supererogation, which is to reject the claim that such acts exist. The advantage of this response is clear: it does not require the consequentialists to compromise their position. They can maintain that it is consequences alone that matter morally and so the right thing to do is whatever brings about the best consequences. The disadvantage of this view though is just as clear: denying the existence of supererogation denies a recognizable (p. 274) feature of our moral lives and will strike many as unacceptably counterintuitive. This problem, we might think, only gets worse if we respond by seeking to eliminate our other deontic concepts as well. This response does well in terms of theoretical purity, then, but fares poorly in terms of accommodating important features of our practices.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

4. Satisficing Consequentialism

How else might consequentialists respond to the problem of supererogation? The obvious approach to pursue next is to look to ways of reformulating consequentialism so that it is compatible with the existence of acts of supererogation. The remainder of this chapter will investigate several such proposals.

The first proposal I will examine is satisficing consequentialism. Michael Slote (1985) proposes a form of consequentialism according to which it is morally acceptable to act in a way that is merely "good enough" rather than performing the best act available. Good enough here should be understood in relation to the consequences of the other available actions (Slote 1984, 155). This gives us the following account:

> *Satisficing Consequentialism*: It is permissible for an agent A to perform act φ at time *t* if and only if the consequences of φ-ing are sufficiently good in comparison to the consequences of the other acts that A could perform at *t*.

Slote argues that this view looks plausible when we consider the following example:

> A warrior has fought meritoriously and died in a good cause, and the gods wish to grant him a single wish for those he leaves behind, before he enters Paradise and ceases to be concerned with his previous life. Presented with such an opportunity, may not the warrior wish for his family to be comfortably well off forever after? And will we from a common-sense standpoint consider him to have acted wrongly or non-benevolently towards his family because he (presumably knowingly) rejected an expectably better lot for them in favor of what was simply good enough? Surely not. (1984, 150–151)

If we accept that it is acceptable for the warrior to choose a "good enough" option here, then this suggests that there is no need to choose the best available act; it is enough to choose an option that is merely "good enough."

This view has a clear advantage over maximizing consequentialism when it comes to accommodating supererogation. As Slote (1985, 53) points out, if we accept that a moral agent is only morally required to do what is "good enough," then this opens up space for the possibility of acts that are morally better than the minimum that is required. If Act A is the best act available and Act B merely "good enough," then act A will be better than the minimum that is required. On a satisficing view, then, there is nothing puzzling about the existence of supererogatory acts.

(p. 275) Moreover, Slote takes this view to follow from how we should think about practical reasons more generally. Suppose, for example, someone is selling her house and decides in advance what an acceptable offer would be based on what she paid for it, what her new house will cost, and what similar houses in the area are selling for. According to Slote, it is perfectly rational for this agent to accept the first offer that meets this amount. There is no requirement of practical reason that the agent seeks to obtain the most mon-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ey possible in this situation. It is important to point out that on Slote's (1985, 38) view this is the case even if the agent knows that she could get a better offer. While some might be tempted to accept a satisficing strategy as a rational decision procedure, Slote's view is more radical. Slote defends satisficing as a criterion of the rightness of actions. This means that even when the agent knows that a better option is available, she is perfectly justified in choosing an option that is merely "good enough."

Such a view faces an obvious question: how can we decide whether the consequences of an action are good enough in comparison to the alternatives? In the absence of a decision procedure here, the view may be seen as incapable of guiding our actions. Setting this worry to one side, the more important problem for our purposes is the problematic counter-examples the view faces. Consider the following case, given by Tim Mulgan:

> *The Magic Game:* Achilles is locked in a room, with a single door. In front of him is a computer screen, with a number on it (call it n), and a numerical keypad. Achilles knows that n is the number of people who are living below the poverty line. He also knows that, as soon as he enters a number into the computer, that any people will be raised above the poverty line (at no cost to Achilles) and the door will open. There is no other way of opening the door. Because of the mechanics of the machine, any door-opening number takes as much time and effort to enter (negligible) as any other.
>
> Achilles enters a number (p) which, although fairly large, is significantly less than n. We ask him why he opted not to raise a further n-p people above the poverty line. He replies that he is a Satisficing Consequentialist who thinks that saving p people from poverty in one day is "good enough." He thus sees no reason to save more people, and doesn't think he's done anything wrong. (1993, 125)

According to Slote's view, there will be some value of p that would make it morally permissible to choose p over n. If this is the case, then choosing n is supererogatory, as it is morally better than the minimally permissible option. This is a strange result, given that it would have been just as easy to save the larger number of people. There seems little justification for Achilles to choose the lower number.[6]

Another problem with ethical satisficing is that it rests on controversial commitments about the nature of practical reasoning. It is fair to say that many are far from convinced by Slote's claim that deliberately choosing a "good enough" option when a better option is available at no cost whatsoever is rationally justified. For instance, Philip Pettit (1984, (p. 276) 12) argues that when an agent is choosing between two available options and evaluates one better than the other, then claiming that the lesser option is "good enough" is no justification for choosing it when a better option is available. Pettit is far from alone in finding rational satisficing puzzling or incoherent.[7] Given that rational satisficing is such a controversial view of practical reasoning, it may well seem preferable for consequentialists to find a different way to respond to the problem of supererogation.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Of course, there is no need for a satisficing consequentialist to commit herself to satisficing more generally. It is perfectly coherent to claim that in morality it is permissible to do only what is "good enough" while when it comes to practical reason more generally agents ought to maximize. Nevertheless, without the appeal to rational satisficing, we are left without a clear rationale for accepting satisficing in the moral domain beyond the fact that the view is compatible with the existence of acts of supererogation.[8]

In summary, satisficing consequentialism does a good job of accommodating supererogation and making sense of this aspect of our moral practices. However, it does this by appealing to a controversial view of practical rationality that many reject. It could be endorsed without appealing to this view, but this is theoretically unsatisfying, as we are left without a clear rationale for accepting a satisficing version of consequentialism.

5. Dual-Ranking Consequentialism

So far we have examined two consequentialist responses to the problem of supererogation. Rejecting supererogation has theoretical appeal as it allows the consequentialist to hold on to the most straightforward version of the view. However, in doing so it denies a familiar feature of moral lives and so is likely to strike many as unacceptably counter-intuitive. Satisficing consequentialism, on the other hand, can accommodate this feature of our moral lives, but it does so in a theoretically unsatisfying way. Where should a consequentialist look next for a solution to the problem? The ideal would be to find a theoretically satisfying way of accommodating supererogation. In this section we will examine an approach that aims to achieve this ideal: dual-ranking consequentialism.

The starting thought behind this approach is that there is a limit to the extent to which morality can demand that we sacrifice our own self-interest in order to promote what would be good from an impartial point of view. As Samuel Scheffler puts the point:

> It is a basic tenet of our commonsense moral outlook that we are justified in devoting some disproportionate degree of attention to our own basic interests, where these are construed as including our fundamental human needs as well as the major activities and commitments around which our lives are organized. (1992, 122)

(p. 277) There is a long history of this thought being used to ground an objection against consequentialism.[9] If consequentialism demands that we perform the act that would bring about the best consequences, then there seems little room for prioritizing our own self-interested concerns, unless we are in the fortunate position of these coinciding with what would be morally best.

However, dual-ranking consequentialists argue that the right response to this thought is not to reject consequentialism but to reform it. We can do so by holding that there are two independent scales by which to evaluate actions. The first evaluates acts from the point of view of moral reasons and the second from the point of view of the other reasons that an agent has for acting. We could then hold that the moral permissibility of an action

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

is not determined by moral reasons alone but is also partially determined by an agent's nonmoral reasons. Portmore has proposed the following version of this view (though as we shall see he no longer holds this position):

> *Dual-Ranking Act Consequentialism:* S's performing *x* is morally permissible if and only if, and because, there is no available act alternative that would produce an outcome that S has both more moral reason and more reason, all things considered, to want to obtain than to want *x*'s outcome to obtain. (2011, 118)[10]

In other words, the moral permissibility of an act is not fully determined by moral reasons. An act A may be the morally best act available, but if the best act all things considered (taking an agent's moral and nonmoral reasons into account) is act B, then the agent has a choice about whether to perform A or B.

Why think though that moral permissibility could be partially determined by some reasons that do not count as moral reasons? Portmore (2011, 123) responds to this question by making the following distinction:

> *Moral Reasons:* Reasons that, morally speaking, count in favor of, or against, performing some action.
>
> *Morally Relevant Reasons:* Any reason that is relevant to determining an act's moral status.

This distinction opens up the possibility that reasons may serve to morally justify a course of action without morally favoring that act. As Portmore (2011, 122) points out, if nonmoral reasons are capable of playing this moral justificatory role, then this means that moral reasons are not morally overriding. On this view moral reasons do not fully determine the moral status of acts. Portmore supports this claim by appealing to the following cases:

> *Fiona's Choice Version 1:* Fiona is about to transfer the balance of her savings account to her current account. She must do this if she is to purchase a new home, and she can do this simply by clicking on a button TRANSFER. Alternatively, if she clicks on (p. 278) a button DONATE, her savings will be transferred, to Oxfam, providing various strangers in the Third World with some considerable benefit. (Paraphrased from Portmore 2011, 125)

Portmore argues that those who want to deny that Fiona has an obligation in this case should accept that this is because the nonmoral reasons are playing a morally justifying role. This can be seen clearly by comparing it to the following case:

> *Fiona's Choice Version 2:* Fiona is in the same position as before except, in this case, if she clicks the button she will lose no money of her own. Instead a philanthropist has agreed to transfer an equivalent sum of his own money to Oxfam. Whichever button she presses, Fiona will purchase her new home, but, by clicking

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> on DONATE, she will also secure a considerable benefit for various others. (Paraphrased from Portmore 2011, 125)

In this case Portmore claims it seems reasonable to think that Fiona is obliged to press DONATE rather than TRANSFER. The reason for the difference is that there are no nonmoral reasons that count against pressing DONATE.

Once we accept that nonmoral reasons are capable of playing this moral justificatory role, then providing an explanation for supererogation is simple (Portmore 2011, 131–136). Suppose an agent has to choose between two available acts, A and B. The moral reasons speak in favor of performing act A, while the nonmoral reasons speak in favor of performing act B. Now we can see that if the nonmoral reasons are of sufficient moral justificatory force, then act A will not be required but will be morally better than another permissible act, act B. We can see, then, that dual-ranking consequentialism can successfully explain the possibility of acts of supererogation. This view has a clear advantage over maximizing consequentialism, as it is able to accommodate the existence of acts of supererogation.

Dual-ranking consequentialism also manages to accommodate the existence of acts of supererogation in a more theoretically satisfying way than satisficing consequentialism. While the theoretical justification for satisficing consequentialism relies on a controversial view of practical rationality, dual-ranking consequentialism presents an intuitively compelling picture of what makes an act supererogatory rather than obligatory. As we have seen, one thought that seems to motivate those who seek to account for the supererogatory is that unless we make room for these acts we risk presenting an unattractive picture of the good life in which morality takes precedence over all other concerns. Dual-ranking consequentialism fits nicely with this thought, as these are acts that would have been obligatory were it not for the nonmoral reasons that count against their performance.

However, it should be noted that dual-ranking consequentialism is somewhat less theoretically satisfying than maximizing consequentialism. As Portmore (2011, 119) concedes, “the move from a single-ranking to a dual-ranking structure is not motivated by any axiological intuitions—that is not motivated [by] our intuitions about which outcomes agents ought to prefer.” Consequentialists must decide whether this loss in (p. 279) theoretical purity is made up for by the ability to accommodate an important feature of our moral practice.

Another worry that has been raised against dual-ranking consequentialism is that it can only accommodate acts of supererogation that would have been obligatory were it not for the levels of sacrifice involved for the agent. We might worry that not all cases of supererogation are like this. This point is made by Ferry (2013, 580), who points out that there may be plausible cases of supererogation that are not opposed by nonmoral reasons. Ferry supports this with the following example:

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> *Gift for Friend:* You see a book on sale and decide to buy it for a friend. If you buy the book it will bring joy to your friend and the pleasure of giving an unexpected gift will also bring joy to you. (Paraphrased from Ferry 2013, 580)

The reason that Ferry takes this example to be problematic for Portmore's account is that this seems to be a case where both the moral and nonmoral reasons speak in favor of performing the act. As a result, it looks as if dual-ranking consequentialism would have to class this act as morally required.

A related problem for dual-ranking consequentialism concerns cases where an agent must make a choice between two trivially different options where one option is slightly better than the other. Suikkanen (2014, 285–286) raises this objection with an example of a couple who must choose between watching one of two television shows. They enjoy both, but they would each get a little more pleasure from watching one of them rather than the other.[11] Again, dual-ranking consequentialism is committed to holding that it is morally required for both people to choose the show that will bring about more pleasure, as this will be both morally preferable (more pleasure overall) and preferable from the point of view of self-interest.

Portmore's (2011) response to these objections is to abandon dual-ranking consequentialism in favor of what he calls common-sense consequentialism. There are many more elements to this view than I have the space to do justice to here. The most important difference for my purposes is that with this view Portmore (2011, 135) accepts the existence of moral reasons that have favoring or enticing strength but no morally requiring strength.[12] This allows a simple response to Suikkanen's example, as the reasons in favor of watching a particular show may be enticing reasons and so would not generate a moral requirement to watch a particular show. It is not obvious, however, how a consequentialist could explain the existence of two different kinds of moral reason, some which have requiring strength and some that do not. One account that has been offered for the existence of two different kinds of moral reasons is offered by Jamie Dreier (2004). He suggests that there may be more than one moral point of view. Reasons stemming from justice may have requiring force while those based in beneficence may not (p. 280) (Dreier 2004, 149).[13] While this response makes sense if we accept a pluralist approach to morality, it is not clear how a consequentialist could reconcile such a view with her commitment to the moral status of actions being fully determined by their consequences. This is not to say that such a view is incompatible with consequentialism but a consequentialist wishing to accept the existence of two different kinds of moral reasons certainly needs to provide some kind of consequentialism-friendly explanation for this divide.

Another objection to the dual-ranking consequentialist solution to the problem is based upon the testimony of those who perform acts of supererogation, who often claim that they would have been unable to live with themselves if they had not acted as they did. Many also report feeling a rewarding sense of inner satisfaction after performing these actions (Archer 2016b). If those who acted in this way would have been unable to forgive themselves had they acted differently, then we might think that acting in this way was not

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

just what the agent had most moral reason to do but also what she had most reason all things considered to do. This is problematic for dual-ranking consequentialism, as it is committed to saying that if this were the case then these acts are required for these agents. This would be an odd result, as it would in effect be committed to claiming that these acts would be required from these agents but would not be required for agents who would not feel satisfied if they acted in this way or guilty if they acted differently. As I have put the point elsewhere, "Effectively, then, the other's less developed moral conscience gets her off the hook from these more demanding obligations" (Archer 2016b, 345). This need not be a devastating objection, as there are those who embrace this conclusion (e.g., Flescher 2003, 115; and Dougherty 2017).[14] Nevertheless, in the absence of an explanation for this, it does seem like a strange upshot of the view.

To sum up, dual-ranking consequentialism manages to accommodate the existence of acts of supererogation in a more theoretically satisfying way than satisficing consequentialism. However, it remains a less theoretically satisfying account of this connection than the straightforward account provided by maximizing consequentialism. In addition, dual-ranking consequentialism is committed to the claim that all acts of supererogation involve agential sacrifice, which faces a number of apparent counter-examples.

6. Reinterpreting Supererogation

Dale Dorsey (2016) provides an alternative consequentialist response to the problem of supererogation. Dorsey's solution to the problem involves rejecting the definition of "supererogation" that I outlined in section 1 and which all of the accounts we have looked at so far have accepted. Instead of seeing supererogatory acts as *morally* optional **(p. 281)** and morally better than another *morally* permissible action, Dorsey claims that supererogatory acts are *rationally* optional and morally better another *rationally* permissible action. This means that supererogatory acts are those that meet the following conditions:

> *Permissible not Required II*: If an act φ is supererogatory, φ is rationally permissible, but is not rationally required.
>
> *Morally Good II*: If an act φ is supererogatory, φ is *especially* morally good or meritorious in comparison to other rationally permissible actions. (Dorsey 2016, 127)

In order to understand Dorsey's proposal we must first examine what Dorsey means by "rational requirement." Dorsey explains his use of the term in the following way:

> [T]here are many different sorts of requirements—not just moral—that I face. I face legal requirements, prudential requirements, requirements of etiquette, requirements of my neighbourhood association. Sometimes these requirements will conflict. But in cases of conflict, it seems natural to ask ourselves what we ought to do really, or all-things-considered. More generally, in the case of conflicting requirements, how should I live? For the sake of brevity, I will refer to this "all-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> things-considered" requirement, which is distinct from, e.g., moral, legal or prudential requirements, as the "rational" requirement, or rational "ought." (2013, 369)

By rational requirements, then, Dorsey is referring to all-things-considered normative requirements.

An advantage of this view is that it is compatible with, though does not entail, a maximizing consequentialist account of moral requirements (Dorsey 2016, 127). Dorsey's account of supererogation allows us to say both that acts of supererogation exist and that *morally* obligatory acts are those that are best supported by moral reasons. It is the *rationally* obligatory acts that can be morally surpassed. This account, then, can retain the theoretically satisfying way in which maximizing consequentialists outline the connection between moral reasons and moral obligations. Dorsey also claims that this response to the problem does better than its rivals at handling cases of supererogation that would have been obligatory were it not for the fact that they require a nontrivial sacrifice on the part of the agent.[15]

Moreover, unlike maximizing consequentialism, Dorsey's account provides us with a way of accommodating the intuitive appeal of the claim that acts of supererogation exist. Dorsey's view allows us to say that Urmson's soldier went beyond the call of duty. It is just that the duty being referred to here is not a *moral* duty but an *all-things-considered normative* duty (2016, 128). On initial appearances, then, it appears that (p. 282) Dorsey's solution offers us a way of capturing both the theoretical appeal of maximizing consequentialism and the intuitive appeal of the claim that acts of supererogation exist.

One objection that Dorsey (2016, 129) considers against his view is that it doesn't really accommodate the intuitive appeal of the claim that acts of supererogation exist. Instead, it simply identifies some other class of acts and shows that they exist. In response, Dorsey claims that while he takes himself to have provided a plausible analysis of supererogation, it is not a major problem for his view if his account is not seen as a plausible view of the concept. In this case he can simply claim to have provided an analysis of another concept, which he terms the "superdupererogatory," which is compatible with his view and that captures everything that an account of supererogation should capture, namely: "actions about which we would say 'hey, you didn't have to do that;' 'that went above and beyond;' and so forth" (Dorsey 2016, 129).

However, as I have argued elsewhere (Archer 2016c, 186), while Dorsey's account can make room for the existence of such acts, it does not do so in a way that fits with the way in which this phrase is used in our ordinary normative discourse. Admittedly, it is plausible to think that the concept of a rational requirement is a recognizable part of our normative discourse. It seems this concept is being appealed to when people make utterances like "You must go to the dentist" or "You must eat more healthily." However, it is implausible to claim that it is this concept that is being picked out by the term "duty." This can be seen if we substitute "have a duty to" for "must" in the previous utterances to read: "You have a duty to do your homework" and "You have a duty to take your medi-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

cine." This substitution seems to change the meaning between the two sentences, at least if we took the original utterances to refer to rational requirements. The new utterances suggest a moral requirement, not a rational requirement. This means that Dorsey's account does not provide a plausible analysis of what is meant by the ordinary language phrase "beyond the call of duty." This is not in itself a devastating objection, particularly not to his main aim of defending an anti-rationalist account of morality's authority. However, it does give us reason to question how successful this account really is in capturing the intuitions that push us toward accepting the claim that acts of supererogation exist.

7. Indirect Consequentialism

I have considered three ways in which consequentialists have sought to provide an account of the connection between the moral requirements and moral reasons that is able to make room for acts of supererogation. All of these accounts have shared the view that consequentialists should be directly linked to the consequences that will be brought about by the various acts an agent could perform at a particular time. In other words, they have all accepted that moral requirements are directly linked to the moral reasons an agent has to perform an act. I will now consider an alternative way in which consequentialists have sought to explain the connection between moral reasons and moral requirements: indirect consequentialism.

(p. 283) Before doing so, it is worth explaining the difference between direct and indirect consequentialism. Walter Sinnott-Armstrong (2015) explains the distinction in the following way: "A direct consequentialist holds that the moral qualities of something depend only on the consequences of that very thing. […] In contrast, an indirect consequentialist holds that the moral qualities of something depend on the consequences of something else." A direct act consequentialism then holds that the moral permissibility of an act depends on the consequences of performing that act. An indirect consequentialist, on the other hand, holds that the permissibility of performing an act depends not on the consequences of performing that act but on the consequences of something else. The obvious question this prompts is: what could this something else be?

The most popular answer to this question is the one given by rule consequentialists (see Hooker, Chapter 23, this volume). According to rule consequentialism, the moral permissibility of an act is determined by whether or not the act is permitted by the set of rules that would bring about the best consequences. An act's moral permissibility is not determined by the consequences of an act but by the consequences of a set of rules being generally accepted. Brad Hooker, for example, defends the following form of rule consequentialism:

> *Rule Consequentialism:* An act is wrong if it is forbidden by the code of rules whose internalization by the overwhelming majority of everyone everywhere in each new generation has maximum expected value in terms of well-being (with

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> some priority for the worst off). The calculation of a code's expected value includes all the costs of getting the code internalized. (2000, 32)[16]

This view can solve the problem of supererogation if we accept that the optimal set of rules will not demand that people always perform the act with the best consequences. Given that such a demanding set of rules might have high internalization costs (Hooker 2000, 78), it is reasonable to think that an alternative code would have better consequences. It is also reasonable to think that an optimal set of rules would also make it permissible and commendable, at least in some circumstances, to perform an act that would have better consequences than the act that is required. This means that rule consequentialism can make room for acts that are morally better than the minimum that morality demands.

Another indirect consequentialist approach is suggested by Ferry (2013), who begins by noting that in *Utilitarianism* John Stuart Mill (2001, 48–49) claimed that the concept of moral wrongness is conceptually tied to sanction. When we say that an action is morally wrong, what we mean is that the agent can be compelled to perform it and ought to be punished in some way for failure to do so. Ferry builds upon this understanding of moral wrongness to offer the following solution to the problem:

(p. 284)

> An act is better or worse by virtue of the reasons for and against it, and in the end, we should accept that we ought to do what we have the most reason to do—we ought to do our best. But the line of duty is determined not by whether we ought to perform the act but by whether we ought to be held accountable for its performance. There will be cases then in which one's obligations involve doing somewhat less than she could do even if she really ought to do her very best. (Ferry 2013, 586)

The key to Ferry's solution is that it is the agent's reasons that determine what it would be morally good or bad to do. Whether or not an act is morally required, on the other hand, is fully determined by the reasons people have to hold the agent accountable.

However, by moving away from the agent's reasons to reasons that others have to react to the agent in certain ways, this view opens up the possibility that reasons that have no impact whatsoever on whether or not the agent should perform the act could change whether or not the act is permissible. For example, suppose someone is walking past a burning building and hears screams for help coming from inside. It seems reasonable to think that it would be supererogatory for him to run in and save the person trapped inside, providing of course that he is not a trained firefighter and the fire is sufficiently dangerous. Suppose, however, that an evil demon tells all of the members of the moral community that she will destroy a whole city of people if they don't require or hold the passerby accountable for saving the woman from the burning building. In this case the reasons concerning whether the passerby should perform the act are more or less unchanged (though perhaps the negative reactions he would face were he not to perform

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

the act provide some additional reason to act in this way). However, despite the agent's reasons being unchanged, the fact that the reasons to hold him accountable are now strongly in favor of doing so means that he now has an obligation to save the woman from the building. It seems odd that a consequentialist would endorse a view that held that whether or not an agent's action is morally required could depend on factors completely unconnected from the consequences of that action.

Even more problematic are cases where an act is prevented from being obligatory by this kind of reason. Suppose that Jane has an obligation to pay her taxes. If the evil demon were to demand that people require Jane not to pay her taxes, then this seems to make it wrong for Jane to pay them. However, it would not make any difference to what Jane has moral reason to do.

The problem is that there can be many reasons to hold someone accountable that are not the kind of reasons that are capable of influencing the deontic status of the action. This is not necessarily a devastating objection to the view. If defenders of the Ferry's approach can give an account of what the right kind of reasons to hold someone to account are, then they will be able to maintain that an act is obligatory if there is most reason *of the right kind* to respond in certain ways to the agent's performance or nonperformance. The challenge for a consequentialist form of this account would be to do so in a way that stays true to the core consequentialist intuition that the consequences of an action are the only features of an action that matter morally.[17]

(p. 285) McElwee offers an alternative indirect consequentialist approach. According to McElwee (2010b), consequentialism provides a plausible account of moral reasons. However, we should not accept the maximizing consequentialist account of moral obligation. Instead, he proposes that consequentialists can offer an alternative account of obligation as determined by norms governing the fittingness of guilt and blameworthiness (McElwee 2010b). While this account would not involve a direct connection between moral permissibility and an action's consequences, McElwee argues that it could allow consequences to play an indirect role in shaping permissibility. It can do so through morally evaluating our moral norms and practices and calling for gradual reform to these norms and the cultivation of our moral sentiments in areas of life where they do not play a large enough role.

This account is also well placed to respond to the problem of supererogation. As our norms for guilt and blameworthiness are less demanding than maximizing consequentialism, there is room on this account for the existence of acts that would have better consequences than alternative permissible actions. In other words, there is room for acts of supererogation.

However, this account also faces objections of its own. One objection that can be raised against it is that there is no one system of moral norms present in the world at any given time. These norms vary significantly across cultures and within the same culture at different points in history. McElwee (2010b, 409) accepts that these changing cultural expectations do indeed change what our moral requirements are. However, there is a further problem here, which is that these social expectations also vary within a culture at a par-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ticular time. How should we decide which group's expectations determine whether or not someone has acted in a morally permissible way? Or will judgments of moral permissibility always have to be indexed to one particular set of social norms on this account? These are questions that will need to be answered satisfactorily in order for this account to be convincing.

Perhaps there are convincing answers McElwee can give to all these questions. Nevertheless, it is worth noting that his account, as with Hooker's and Ferry's, moves us away from the core consequentialist idea that an action's moral status is determined by its consequences. Consequentialists must decide whether this is a price worth paying in order to secure a solution to the problem of supererogation.

Before concluding this section, though, it is worth briefly mentioning a number of reasons that have been given for thinking that a more indirect link between consequences and an act's deontic status may be preferable from a consequentialist point of view. The first suggestion is made by Urmson (1958, 70), who argues that a set of moral norms that demands too much from people would erode the force of moral address and could lead to those norms having less motivational force, as people become less sensitive to these forms of moral criticism.

Similarly, Claire Benn (2018) argues that an overly demanding approach to morality faces similar problems to an overly demanding approach to practical standards more generally. Perfectionism, holding oneself to an overly high standard and being overly critical of one's behavior, is often counter-productive, leading people to per (p. 286) form their tasks less effectively than they would have if they had held themselves to lower standards. According to Benn, the same is true in the moral domain: holding people to highly demanding moral standards may lead people to perform actions that are morally worse than the actions they would have performed had they been held to lower standards. Both arguments push us toward the thought that the overall consequences of a less demanding set of moral standards may be better than those of a more demanding set.

In addition, Suikkanen (2014, 287) argues that the availability of alternative permissible options can increase the value of choosing a particular option. First, we value having free choice (or at least many of us do), so having these options will be instrumentally valuable in that in gives us something we desire or find pleasurable (Suikkanen 2014, 288). Another way in which being able to choose from different permissible options is instrumentally valuable is that it can give us the opportunity to learn something about ourselves (Suikkanen 2014, 288). According to Suikkanen, the availability of options also has constitutive value, in that they constitute parts of a larger whole that is intrinsically valuable. First, they may have representative value (Suikkanen 2014, 289). The fact that the agent chose this act over other options may represent the agent's inner life. They may also have symbolic value as a being part of a complex whole that recognizes our rational capacities (Suikkanen 2014, 290). All of these arguments suggest that from a consequentialist point of view, the world in which an agent is morally permitted to choose a suboptimal option can be better than one in which she is not permitted to choose this option.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

8. Concluding Remarks

In this chapter I have sought to provide an overview of the various ways in which consequentialists have responded to the problem of supererogation. As we have seen, there are two general classes of response that can be made. First, consequentialists can hold firm to the theoretical purity of maximizing act consequentialism and deny the existence of acts of supererogation. Alternatively, consequentialists can seek to provide an alternative account of the relationship between moral reason and moral obligation that is able to accommodate the existence of acts of supererogation. While each of these alternative accounts faces unique problems, they are also all vulnerable to the objection that the ability of these accounts to accommodate intuitions about supererogation comes at the cost of providing a less straightforward articulation of the core consequentialist idea that an act's moral status is determined by its consequences. This may seem to be something of an impasse, where the theoretical considerations that favor maximizing act consequentialism clash with the intuitive appeal of the existence of acts of supererogation.

Given this situation, I suggest that in order to make progress on this debate, consequentialists should focus their attention on the question of whether the concept of moral (p. 287) obligation has a useful role to play in their moral theory and, if so, what role this is (a discussion already begun in the debate between Norcross, on the one hand, and Lang and McElwee, on the other). Unless this concept has any useful role to play in consequentialist moral theory, then the whole discussion concerning how consequentialists can accommodate supererogation can be abandoned. If it does have a useful role to play, then the discussion of how consequentialists can accommodate supererogation should be guided by the role that moral obligation should play in their theory.

References

Archer, Alfred. 2016a. "Moral Obligation, Self-Interest and the Transitivity Problem." *Utilitas* 28, no. 4: 441–464. doi:10.1017/S0953820816000091

Archer, Alfred. 2016b. "Supererogation, Sacrifice, and the Limits of Duty." *Southern Journal of Philosophy* 54, no. 3: 333–354. doi:10.1111/sjp.12176

Archer, Alfred 2016c. "The Supererogatory and How Not To Accommodate It." *Utilitas* 28, no. 2: 179–188. doi:10.1017/S095382081500032

Archer, Alfred 2018. "Supererogation." *Philosophy Compass* 13, no. 3.

Benn, Claire (2018). "The Enemy of the Good: Supererogation and Requiring Perfection." *Utilitas* 30, no. 3: 333–354.

Bradley, Ben. 2006. "Against Satisficing Consequentialism." *Utilitas* 18, no. 2: 97–108.

Byron, Michael. 1998. "Satisficing and Optimality." *Ethics* 109, no. 1: 67–93.

Chappell, R. Y. 2019. "Willpower Satisficing." *Noûs* 53, no. 2: 251–265.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Crisp, R. 2006. *Reasons and the Good*. Oxford: Oxford University Press.

De Lazari-Radek, K., and Singer, P. 2014. *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. Oxford: Oxford University Press.

Dorsey, D. 2013. "The Supererogatory, and How to Accommodate It." *Utilitas* 25, no. 3: 355–382.

Dorsey, D. 2016. *The Limits of Moral Authority*. Oxford: Oxford University Press.

Dougherty, T. 2017. "Altruism and Ambition in the Dynamic Moral Life." *Australasian Journal of Philosophy* 94, no. 4: 716–729. doi:10.1080/00048402.2016.1256331

Dreier, J. 2004. "Why Ethical Satisficing Makes Sense and Rational Satisficing Does Not." In *Satisficing and Maximising*, edited by Michael Byron, 131–154. Cambridge: Cambridge University Press.

Feldman, F. 1986. *Doing the Best We Can*. Dordrecht, the Netherlands: Reidel.

Ferry, M. 2013. "Does Morality Demand Our Very Best? Moral Prescriptions and the Line of Duty." *Philosophical Studies* 165, no. 2: 573–589.

Flescher, A. M. 2003. *Heroes, Saints and Ordinary morality*. Washington, DC: Georgetown University Press.

Harwood, S. 2003. "Eleven Objections to Utilitarianism." In *Moral Philosophy: A Reader*, edited by Louis P. Pojman, 3rd ed., 179–192. Indianapolis: Hackett.

Heyd, David. 1982. *Supererogation: Its Status in Ethical Theory*. Cambridge: Cambridge University Press.

Hooker, Brad. 2000. *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*. Oxford: Oxford University Press.

Horgan, T., and Timmons, M. 2010. "Untying a Knot from the Inside Out: Reflections on the "Paradox' of Supererogation." *Social Philosophy and Policy* 27: 29–63.

(p. 288) Kagan, Shelly. 1984. "Does Consequentialism Demand Too Much? Recent Work on the Limits of Obligation." *Philosophy and Public Affairs* 13, no. 3: 239–254.

Kagan, Shelly. 1989. *The Limits of Morality*. Oxford: Clarendon Press.

Kawall, J. 2003. "Self-Regarding Supererogatory Acts." *Journal of Social Philosophy* 34, no. 3: 487–498.

Kotarbinski, T. 1914. "Utilitarianism and the Ethics of Pity." *Nowe Tory* (i-ii). Translated by W. Rabinowicz (2000) and reprinted in *Utilitas* 12, no. 1: 80–84.

Lang, Gerald. 2013. "Should Utilitarianism Be Scalar?" *Utilitas* 25, no. 1: 80–95.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

McElwee, Brian. 2010a. “Should We De-moralize Ethical Theory?” *Ratio* 23, no. 3: 308–321.

McElwee, Brian. 2010b. “The Rights and Wrongs of Consequentialism.” *Philosophical Studies* 151, no. 3: 393–412.

McNamara, Paul. 1996. “Making Room for Going beyond the Call.” *Mind* 105, no. 419: 415–450.

Mill, J. S. 2001. *Utilitarianism*. Indianapolis: Hackett.

Mulgan, Tim. 1993. “Slote’s Satisfying Consequentialism.” *Ratio* 6: 121–134.

Norcross, Alastair. 1997. “Good and Bad Actions.” *Philosophical Review* 106, no. 1: 1–34.

Norcross, Alastair. 2006. “Reasons without Demands: Rethinking Rightness.” In *Contemporary Debates in Moral Theory*, edited by James Lawrence Dreier, 6–38. Malden, MA: Blackwell.

Pettit, Philip. 1984. “Satisficing Consequentialism.” *Proceedings of the Aristotelian Society, Supplementary Volumes* 58: 165–176.

Portmore, D. 2011. *Commonsense Consequentialism: Wherein Morality Meets Rationality* Oxford: Oxford University Press.

Portmore, D. W. 2017. “Transitivity, Moral Latitude, and Supererogation.” *Utilitas* 29, no. 3: 286–298. doi:10.1017/S0953820816000364

Rabinowicz, W. 2000. “Kotarbinski’s Early Criticism of Utilitarianism.” *Utilitas* 12, no. 1: 79–80.

Schleffer, S. 1992. *Human Morality.* New York: Oxford University Press.

Sider, T. 1993. “Asymmetry and Self-Sacrifice.” *Philosophical Studies* 70: 117–132.

Sidgwick, H. (1907)1981. *The Methods of Ethics.* 7th ed. Indianapolis: Hackett.

Sinnott-Armstrong, Walter. 2015. “Consequentialism.” In *The Stanford Encyclopedia of Philosophy* (Winter 2015 edition), edited by Edward N. Zalta. **https://plato.stanford.edu/archives/win2015/entries/consequentialism/**.

Slote, M. 1984. “Satisficing Consequentialism.” *Proceedings of the Aristotelian Society, Supplementary Volumes* 58: 139–176.

Slote, M. 1985. *Common-Sense Morality and Consequentialism.* London: Routledge & Kegan.

Suikkanen, J. 2014. “Consequentialist Options.” *Utilitas* 26, no. 3: 276–302.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Urmson, J. O. 1958. “Saints and Heroes.” Reprinted in *Moral Concepts*, edited by Joel Feinberg (1969). Oxford: Oxford University Press.

Vessel, J. P. 2010. “Supererogation for Utilitarianism.” *American Philosophical Quarterly* 47: 299–317.

Williams, B. 1973. “A Critique of Utilitarianism.” In *Utilitarianism; For and Against*, edited by J. J. C. Smart and B. Williams, 77–150. Cambridge: Cambridge University Press.

Wolf, S. 1982. “Moral Saints.” *Journal of Philosophy* 79, no. 8: 419–439.

Notes:

([1]) While Urmson’s paper opened up the discussion of the concept in modern Western philosophy, his paper was not the first to discuss the concept. As Rabinowicz notes (2000, 79), Kotarbinski published a short note in 1914 in which he argues that utilitarianism is unable to accommodate the supererogatory (1914/2000).

([2]) This claim is endorsed by Ferry (2013), Horgan and Timmons (2010, 37), and Portmore (2011, 91).

([3]) Though not all use the term “morally better,” the following endorse the claim that supererogatory acts are morally better than nonsupererogatory acts: Ferry (2013, 574), Heyd (1982, 5), and Portmore (2011, 92). As McNamara points out, we need to appeal to the concept of “The Minimum That Morality Demands” in order to make sense of this (1996, 427). Harwood (2003) and Vessel (2010, 302) define the supererogatory in terms of betterness for others. This account of moral betterness is easier for maximizing consequentialism to accommodate. However, it cannot accommodate self-regarding acts of supererogation. If we accept the existence of such acts (see Kawall 2003, for a defense), then we cannot understand supererogation in terms of betterness for others.

([4]) Vessel (2010, 302) also considers what he calls “ties at the top” phenomena as a potential consequentialist solution to the problem and rejects it for the same reasons.

([5]) Similarly, Crisp (2006) argues that the important questions in ethics concern what we have reason to do and do not concern moral concepts or properties.

([6]) There are a number of different satisficing positions available, and it is not clear that all are vulnerable to counter-examples such as this. For a critique of several distinct satisficing views, see Bradley (2006).

([7]) See Byron (1998) and Dreier (2004).

([8]) One form of satisficing consequentialism that may be immune to these objections is Chappell’s (2019) willpower-satisficing consequentialism. Unfortunately, I do not have the space to discuss this view adequately here.

([9]) See, for example, Williams (1973) and Wolf (1982).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([10]) See also Sider (1993) and Vessel's (2010, 305) egoistically adjusted utilitarianism.

([11]) Or, for nonhedonists, bring about slightly more well-being, however this is understood.

([12]) See Horgan and Timmons (2010) for a discussion of such reasons in response to the problem of supererogation. This discussion is not explicitly a consequentialist one.

([13]) See Archer (2016a, 460) for further discussion of this point.

([14]) For further discussion, see Archer (2016b, 348–349).

([15]) Dorsey claims that his account also does better than its rivals at handling cases of supererogation that would have been obligatory were it not for the fact that they require a nontrivial sacrifice on the part of the agent. I do not have space to evaluate this claim here, though I have argued against it elsewhere (see Archer 2016a). See also Portmore (2017).

([16]) Hooker's account contains further conditions concerning internalization costs and adjudicating between two sets of rules that are best equal. I omit these to avoid overcomplicating the discussion.

([17]) Ferry (2013: 586–587) offers a number of suggestions of what these reasons are that are not obviously compatible with this core consequentialist idea.

Alfred Archer

Alfred Archer is Assistant Professor of Philosophy at Tilburg University and a member of the Tilburg Center for Logic, Ethics, and Philosophy of Science. His primary research interests are in moral philosophy and moral psychology, particularly supererogation, the nature and ethics of admiration, and the ethics of fame. He also has research interests in applied ethics, political philosophy, and the philosophy of sport. He is currently working on a project investigating the nature, ethics, and value of admiration, funded by an NWO Veni grant. For up-to-date information about his research, visit http://alfredarcher.weebly.com/.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Must I Benefit Myself?

Michael Cholbi

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.18

Abstract and Keywords

Morality seems to require us to attend to the good of others, but it does not require that we assign any importance to our own good. Standard forms of consequentialism thus appear vulnerable to the *compulsory self-benefit objection*: they require agents to benefit themselves when doing so is entailed by the requirement of maximizing overall impersonal good. Attempts to address this objection by appealing to ideally motivated consequentialist agents; by rejecting maximization; by leveraging consequentialist responses to the more familiar special relationships and demandingness objections; or by appealing to dual rankings of moral and all-things-considered reasons fall short of adequately answering this objection. A satisfactory response to the compulsory self-benefit objection is elusive because of consequentialism struggles to account for directed options (in this case, an option not to maximize one's own good but not that of others) and for moral considerations that do not rest on the value of outcomes or states of affairs.

Keywords: self-interest, moral options, moral permissibility, self-sacrifice, directed duties

ALTRUISM—THE selfless concern for others—dominates most popular conceptions of morality. Our moral role models tend to be those with reputations for devotion to others' well-being (such as Mother Theresa) or for great personal sacrifice to serve others (such as Nelson Mandela). On this altruistic picture, benevolence and sympathy are the primary moral virtues. A number of philosophers have agreed, arguing that altruism is all there is to the moral point of view. Morality, on this picture, is "purely and essentially other-regarding."[1]

On its face, consequentialism accords well with this altruistic conception of morality. A central message of consequentialist moral thinking has been that nearly all moral agents are exceedingly partial in their moral outlooks and so cast too narrow a net of moral concern; taking altruism seriously entails becoming far more concerned for the needs or interests of others, including those of distant strangers, future generations, and nonhuman animals, groups often consigned to the periphery of ordinary moral thought. This same message—that morality places stringent demands on our altruism—is also evident in con-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

sequentialist writings on practical ethics, which teach that the affluent have moral obligations to donate large portions of their wealth to assist the global poor[2]; that having met their children's needs, parents ought to prioritize other children's needs above their children's wants[3]; and that societies should implement schemes to allocate scarce goods such as transplant organs with the aim of saving the greatest number of lives overall.[4] Consequentialism thus seems to embrace a picture of morality wherein concern for others is paramount while concern for oneself is minimal or nonexistent, a picture in which morality assigns "*no* positive value to the well-being or happiness of the (p. 254) moral agent of the sort it clearly assigns to the well-being or happiness of *everyone other than the agent*."[5]

It may therefore be surprising that standard versions of consequentialism seem to entail that sometimes individuals not only may, but in fact must, benefit themselves. Consider these two examples:

> *Enrollment*: Josephine, a university philosophy student, is searching for a final course to complete her schedule for the next term. She sees that there is a single spot available in a course offered by her favorite ethics professor. Josephine is about to enroll in the course using the university's online system, but at the last moment, she texts her friend Kelly, who finds ethics tolerable but not engaging, to let her know of the spot in the ethics course, which Kelly takes. Despite knowing that she would benefit more from the ethics course than Kelly would, Josephine ends up enrolling in a lackluster course in modal logic.
>
> *Evacuation:* Igor is a solitary, elderly man with no surviving family and few acquaintances. He is frail and undergoing steady cognitive decline. Working or contributing to socially valuable activities is therefore not a viable possibility for him. These challenges notwithstanding, Igor is able to address his basic physical needs on his own and enjoys a good quality of life. He is content to spend his days reading war histories and watching sports on television. One afternoon, a fire breaks out in his apartment building. Rescue crews quickly arrive and give an immediate evacuation order for all of the apartment's residents. Though he is capable of evacuating the building, Igor remains in his apartment, knowing that he could well die from the fire. He is killed when he inhales the smoke that soon pervades the building.

From a naïve or pretheoretical moral view, how should these examples be analyzed and evaluated? In Enrollment, Josephine acts altruistically inasmuch as she foregoes her slot in the ethics course so that Kelly can enroll in it. This seems morally permissible, even laudable. In Evacuation, we might find Igor's decision odd; we would naturally wonder why, unless he had been depressed, and so on, he would willingly choose to end a life that seems congenial if unambitious. And our moral reaction might differ if ending his life caused others to suffer or to grieve. But given his isolation and his inability to contribute much to the happiness of others, his death affected only him, and while questions might be raised about whether Igor acted *prudently*, he does not seem to act *wrongfully*.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Orthodox forms of consequentialism have difficulty validating such responses though. Standard act consequentialism holds that agents are obligated to bring about the best outcome, that is, the outcome that realizes the greatest good overall in comparison with alternatives. But the agents in Enrollment and Evacuation do not satisfy this standard. In the former, Josephine stood to gain more from enrolling in the ethics course than Kelly did. By giving Kelly the opportunity to enroll in the ethics course, Josephine made Kelly better off, but not by a quantity as great as the quantity of goods Josephine would have enjoyed if she had enrolled in the ethics course herself. By the (p. 255) lights of standard act consequentialism, Josephine's altruistic act was wrong: she could have, but did not, perform the act resulting in the greatest good overall. Let us call examples such as Enrollment, in which an agent's action results in good for others but maximal good would have resulted from her choosing to direct greater benefits to herself instead, *nonoffsetting failures to self-benefit.*

Evacuation is not an example of a nonoffsetting failure to self-benefit. For Igor's not evacuating does not result in some quantity of goods being enjoyed by others. Rather, Igor's choosing not to evacuate amounts to him failing to benefit himself, but no one else benefits thereby either. Evacuation thus represents an instance of *pure failure to self-benefit.* Nevertheless, standard act consequentialism condemns Igor's choice on the same general grounds as it condemns Josephine's: he too failed to bring about the greatest good in comparison with alternatives. Put differently, when only our own well-being is at stake, consequentialism seems to demand that we maximize our well-being, a result at odds with the sense that failing to benefit oneself (or to benefit oneself to the utmost) may be unwise but is rarely immoral.

These examples illustrate that, despite its theoretical and practical emphasis on benefitting others, consequentialism appears susceptible to a *compulsory self-benefit objection*: it requires that individuals direct benefits to themselves when doing so is entailed by the consequentialist demand to maximize good overall, despite its being intuitively morally permissible, or even praiseworthy, for individuals not to direct those benefits to themselves.

The compulsory self-benefit objection has force because it flows directly from features of consequentialism that many of its adherents find attractive. What unites consequentialist theories is the conviction that the moral status of actions is a function solely of the outcomes of those actions, and in particular, how good or desirable these outcomes are in comparison to one another. The compulsory self-benefit objection gets its traction in part from this conviction, since failures to benefit oneself seem to result in outcomes that are worse overall. Likewise, consequentialists generally point with pride at its incorporating a strong notion of impartiality, wherein the concerns, interests, well-being, and so on of all affected by an action are taken into account equally when determining how good that action's outcome is.[6] Josephine and Igor are seemingly required to benefit themselves because doing so assigns their own concerns, interests, well-being, and so on equal weight to that of others. The theoretical interest of the compulsory self-benefit objection there-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

fore resides in the fact that it arises from features of consequentialism that seem to speak in the theory's favor.

Compared to other objections to consequentialism, the compulsory self-benefit objection has garnered little attention. The purpose of this article is therefore to explore how consequentialists might best answer the compulsory self-benefit objection. While some might find "biting the bullet" sufficient—acknowledging the force of the objection but maintaining that the other merits of consequentialism outweigh that force—I shall (p. 256) assume that consequentialists would prefer to answer the objection on its face in a way that does minimal damage to their theory. The challenge, then, is to ascertain how consequentialism might permit individuals not to benefit themselves without relinquishing the commitments that give consequentialism its theoretical appeal.

I cannot consider every strategy by which consequentialists might try to answer the compulsory self-benefit objection. Here I will canvass several strategies that either strike me as particularly promising or have not been previously explored.[7] Ultimately, the strategies I consider here all founder on one or more of the following three worries: first, they do not account for a broad enough spectrum of the intuitively morally permissible instances of failures to self-benefit (for example, pure failures to self-benefit); second, they address the objection through ad hoc maneuvers that license failures to self-benefit without motivating these maneuvers by appeal to consequentialist principles or commitments; or third, they do not adequately explain how the permissibility of failing to benefit oneself is grounded in a moral option—that is, they do not provide compelling explanations of how *both* benefitting oneself and failing to do so are morally permissible. I do not take their failures to demonstrate the impossibility of a credible consequentialist response to the compulsory self-benefit objection. But their failures do suggest that the objection is more formidable than the extant literature implies, and defenders of consequentialism would be wise to exert more effort in explicitly answering it.

Our discussion unfolds as follows. In the next section, I consider whether consequentialist theories that evaluate actions in terms of their motives can answer the compulsory self-benefit objection. Section 2 turns to two ways of addressing the objection by deviating from standard maximizing forms of consequentialism. In sections 3 and 4, I argue that the objection can be classified as an instance of more familiar objections to consequentialism—that it gives insufficient due to special relationships and is exceedingly demanding, respectively—but consequentialist responses to those more familiar objections do not readily apply to the compulsory self-benefit objection. Section 5 considers whether dual-ranking act consequentialism has the resources to answer this objection through its appeal to the distinction between an agent's moral reasons and her all-things-considered reasons for action. In my concluding section, I propose that an adequate consequentialist response to the compulsory self-benefit object must account for options directed at the self, a notion difficult to reconcile with central consequentialist commitments.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

1. Motive Consequentialism

As noted at the outset, altruism and beneficence are generally applauded, greed and self-indulgence generally deplored. This might suggest that a consequentialist theory that (p. 257) takes motives rather than acts as its focus could offer a satisfactory response to the compulsory self-benefit objection.

Suppose that an individual S's action A is obligatory only if S would perform A if S were to act on the basis of a set of desires D that, on balance and over time, leads S to maximize overall goodness through her actions. We might view this as the core of an "ideal motivation" consequentialist theory, wherein acts are judged indirectly, not in terms of whether they themselves maximize overall goodness but in terms of whether an agent whose motivations conduce to maximizing goodness would perform such acts.

A crucial question regarding this theory is, "what is D?" This version of consequentialism would be ineffective against the compulsory self-benefit objection if D is the desire to maximize overall goodness. For an agent motivated by that desire *would* choose to maximize overall goodness, thus ruling out failures of self-benefit. However, some philosophers have argued that the motives the possession of which would maximize overall goodness would not include the motive to maximize overall goodness, much in the way that being motivated to maximize pleasure may actually undercut the aim of maximizing one's pleasure.[8] The question at hand, then, is whether the set of goodness-maximizing desires would include, or allow for, failures to self-benefit. On its face, we might expect that a goodness-maximizing set of desires would be primarily benevolent or other-focused: given our egoistic propensities, a set of desires that counteracts those propensities by largely directing our attention and concern toward the interests of others would lead to high levels of goodness overall. A community of altruists, we would anticipate, would be better off than a community of egoists. If so, then agents motivated by those desires that maximize overall goodness would only rarely benefit themselves when doing so is best overall. An ideal motive consequentialism might therefore give significant breadth to failures to benefit oneself.

This reasoning is too quick though. For one, in cases of pure self-benefit, no one else's interests are at stake, and yet it seems permissible not to maximize benefits to oneself. An ideally motivated agent should be willing to benefit herself in such a case, in opposition to her generally benevolent motives. This reasoning also neglects how the pursuit of self-interest can sometimes redound to the benefit of others. One need not be a dyed-in-the-wool Mandevillean to recognize that at least sometimes the ardent pursuit of one's self-interest can also promote the interests of others. When we pursue our own good through cooperative endeavors with others or when we produce goods we exchange with others, we increase overall well-being despite being guided by selfish motives. Finally, there are dangers in excessive benevolence or concern for others. A theme within some feminist writings is that traditional patriarchal cultures can encourage women to prioritize the interests of others over women's own interests, resulting in the reinforcement of oppressive practices and the diminution of women's self-respect.[9] So while an ideally motivated

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

consequentialist agent might be predominantly altruistic, her motivations can err too far in that direction, and there may be a significant number (p. 258) of instances wherein purely altruistic motivation fails to be best from a consequentialist perspective because one's own interests are neglected. Sorting out just how often an ideally motivated consequentialist agent would attend to her self-interest is a complex empirical matter that I cannot hope to settle here. Still, any such consequentialist theory will face the problem of how to situate the interests of the self within a largely altruistic motivational economy without ignoring the interests of the self in ways that consequentialism rejects.

2. Nonmaximization Strategies

Consequentialists may seek to address the self-benefit objection by adopting a nonmaximizing standard of obligatory action that allows for failures to self-benefit. For example, satisficing consequentialists may propose that so long as agents produce a sufficient amount of good through their actions, they meet their obligations. And in cases of failures of self-benefit, agents may fail to maximize overall good while still producing enough good by consequentialist lights. In both Enrollment and Evacuation, perhaps Josephine and Igor do *well enough* even if they do not do what is impartially best.

Moving to a satisficing rather than maximizing standard has the advantage of explaining how self-benefit is an option: an agent acts rightly if, by failing to benefit herself, she either does well enough or if she exceeds the satisficing standard. A satisficing standard thus permits, but does not require, maximization. The difficulty with this strategy is that the moral permission not to benefit oneself appears very wide, and there is no particular reason to expect that all instances in which an agent fails to benefit herself will meet the satisficing standard. Take Evacuation: suppose Igor foregoes a very large amount of good to himself by failing to evacuate. If doing so is permissible, then satisficing consequentialism would, in order to vindicate this conclusion, have to depart significantly from a maximizing standard. But there is no apparent basis for adopting such a lenient satisficing standard aside from its ability to address failures to self-benefit. Why, after all, should we expect that a qualitative property actions may have (i.e., that they fail to benefit oneself to the greatest degree) will coincide with a quantitative property of actions (i.e., that they produce enough good to meet the satisficing consequentialists' demands) in every instance? We have better reason, I suggest, to expect that any intuitively plausible satisficing standard will at least sometimes disallow apparently permissible failures of self-benefit.

Moreover, such a wide departure from maximization would presumably apply not only in instances of failures of self-benefit but across the board, that is, to situations where an individual falls short of maximization solely with regard to how much good her actions produce for others. But absent some rationale for restricting these deviations only to self-regarding deviations, a satisficing strategy for answering the compulsory self-benefit objection runs the risk of asking too little of agents with respect to others' good.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(p. 259) Another way to answer the compulsory self-benefit objection by jettisoning strict maximization is to incorporate a distinction between benefits to an agent and benefits to others. Ted Sider, for example, has put forth a self/other asymmetrical consequentialist theory wherein if agent S performs act A, then A is obligatory if and only if (1) no other action produces more impartial overall good than A, and (2) no other action produces more good from the "selfless perspective" that excludes those goods that A provided to S.[10] Sider furthermore proposes that if two alternative actions A and B are such that A satisfies condition (1) but not condition (2), while B satisfies condition (2) but not condition (1), then A and B represent a moral "tie," in which case an individual can permissibly perform either A or B. Sider's revised consequentialist standard would appear to answer the compulsory self-benefit objection insofar as it denies that self-benefit is morally obligatory. If we apply it to Enrollment, it would seem true that Josephine's enrolling in the ethics course satisfies condition (1) but not condition (2), since there is no action that results in more impartial good, but there is an action—facilitating Kelly's enrollment—that produces more good when the goods that might accrue to Josephine are excluded. Her facilitating Kelly's enrollment is permissible (though Josephine's enrolling would be permissible as well).

In the case of Evacuation, however, Sider's asymmetrical consequentialism stumbles. Igor's evacuating rather than remaining in his apartment satisfies Sider's condition (1), since his evacuating benefits him and thus (assuming that others' interests are unaffected by his decision) would be better with respect to overall personal good. But Igor's remaining also satisfies Sider's condition (2), because no other action besides Igor's remaining (including his evacuating) produces more good from the "selfless perspective" that excludes whatever goods that Igor gains from remaining. His remaining rather than evacuating makes no difference from that perspective. Thus, remaining versus evacuating is not a tie, and Igor violates his moral obligations by remaining. Sider's revision to consequentialism thus seems to yield intuitively plausible answers in cases of offsetting failures to self-benefit but not in cases of pure failures to self-benefit.

But even if Sider's asymmetrical consequentialism succeeded in making sense of pure failures to self-benefit, it has the deeper theoretical defect of addressing the compulsory self-benefit objection in a largely ad hoc manner. As Sider himself recognizes, his revisions to standard consequentialism may lead to more plausible results about the "moral normative status of actions," but these results are not grounded in any "independently important axiological facts."[11] And consequentialists themselves should have reservations about modifying the consequentialist standard to address the compulsory self-benefit objection. Both satisficing consequentialism and Sider's asymmetrical consequentialism tacitly reject maximization, and the latter rejects impartiality. To the extent then that maximization and impartiality are compelling features of the consequentialist moral framework, these revisions to the consequentialist framework represent theoretical costs, and at least some consequentialists (I expect) will find these costs (p. 260) not a reasonable concession to the theory's critics but the repudiation of what makes consequentialism appealing in the first place.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

3. Special Relationships

Consequentialists have proven very resourceful in answering the many objections to which their theory has been subject. Thus, if the compulsory self-benefit objection resembles an extant objection to which consequentialists have given compelling replies, then an adequate response to the objection may emerge. Let us now consider whether such a strategy may succeed in connection with the complaint that consequentialism places inadequate stock in *special relationships*.

This objection holds that because consequentialists generally insist on impartiality, their theory cannot make sense of instances where partiality is morally permissible, even laudable. Chief among these instances of laudable partiality is our tendency to accord the interests of those close to us—our romantic partners, children, friends, and family members—greater weight in our decision making than the interests of mere strangers.

The compulsory self-benefit objection can be analyzed as an idiosyncratic instance of the special relationships objection. If we are morally permitted not to benefit ourselves even when doing so would be required by the consequentialist demand to maximize overall goodness, then this permission can be viewed as reflecting a permissible "partiality" toward ourselves. This partiality is different from the partiality licensed by other special relationships. For whereas those relationships seem to entitle us to accord certain individuals' interests *greater* weight in our decision making, the special relationship to self seems to license us assigning *lesser* weight to our own interests in our decision making, that is, to allow us not to benefit ourselves. Our distinctive relationships toward particular others establishes special obligations; our distinctive relationship to ourselves establishes special options.

It is unlikely though that the arguments consequentialists have deployed in order to validate special relationships can be applied in the case of compulsory self-benefit and the special relationship to self.

One such argument is that attending to our special relationships, despite being a deviation from impartial consequentialist reasoning, in fact results in the best overall consequences. Realizing impartially best outcomes may sometimes be the result of partiality, so that when we care for our friends to a greater degree than we care for strangers, we strengthen practices that, in general and for the most part, redound to everyone's benefit. It may be, then, that consequentialism does not only permit the cultivation and recognition of special relationships. Rather, the consequentialist project of maximizing overall goodness will *require* their cultivation and recognition. But this argument extends uncomfortably to our special relationship to self: being "partial" by discounting our own well-being may result in greater overall well-being in nonoffsetting failures of (p. 261) self-benefit such as Enrollment, yet it seems unlikely that it results in greater overall well-being in pure failures to self-benefit such as Evacuation. For recall that Igor's remaining in his apartment harms him with no compensating gain to others.[12]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

A second way consequentialists have addressed the special relationships objection is by claiming that such relationships are not means to valuable ends or outcomes, but are intrinsically valuable in their own right. The partiality shown in special relationships is necessary and morally justified because it is part and parcel of something good in itself. Here too it is difficult to see how this line of thought can be extended to the self and apparently permissible failures to benefit ourselves. Again, the failure to benefit ourselves is an option; that is, it is also permissible for agents to choose to benefit themselves as well. To fail in one's special *obligations* to one's friends, say, is to undercut whatever value our friendships have. But in the case of the special relationship to ourselves, the relationship cannot be undercut by what we do (or fail to do) in the way of benefitting ourselves, and if something is good in itself regardless of whether our actions maintain it or not, it is hard to decipher how the putative good is a good in any recognizably consequentialist sense. If an agent could equally maintain this good by benefitting herself and by failing to do so, then this good is not an outcome of what she does and so does not look like a good with which consequentialist morality is concerned.

4. Demandingness

Another familiar objection to consequentialism that resembles the compulsory self-benefit objection is the complaint that consequentialism is too *demanding*.[13]

The usual gloss on this objection is that consequentialism requires us to forego more of our interests or well-being than it is reasonable or morally defensible to demand. The impartial maximization of overall good appears to entail that we are obligated to forego most luxuries in order to donate large sums of money to the poor; that we are obligated to forego the pleasures of meat eating in order to curtail harms to animals; or that we are obligated to choose our careers not on the basis of our aspirations or values but on the basis of which careers will do the most good. Such demands, the objection goes, are unreasonable. Either a more plausible consequentialist theory, making less extensive demands on individuals' well-being, must be expounded or we should reject consequentialism altogether.

(p. 262) So described, the demandingness objection and the compulsory self-benefit objection may seem unrelated. After all, the latter is the complaint that consequentialism requires us to benefit ourselves, whereas the former is the complaint that consequentialism precludes us from benefitting ourselves in ways that seem morally defensible. However, the demandingness objection can be recast in terms of *options*, with the result that the compulsory self-benefit objection is an instance of it: consequentialism is unreasonably demanding in depriving us of options,[14] including the apparently permissible option of failing to benefit ourselves. So depicted, that consequentialism disallows failures to self-benefit shows that it encumbers us not only in terms of our well-being but in terms of our exercising our capacities of choice. Independently from its constraining (perhaps unreasonably) our pursuit of our well-being, consequentialism is too "confining" with respect to the options it leaves us,[15] including options regarding self-benefit.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Debates about the demandingness of consequentialism have generally concerned how to temper the demands of beneficence rather than the demands of self-interest. Yet if the compulsory self-benefit objection is an instance of the demandingness objection, then that provides a reason to suppose that consequentialist rejoinders to the demandingness objection might also serve as effective rejoinder to the compulsory self-benefit objection. In particular, this strategy would be effective against the compulsory self-benefit objection if consequentialist answers to the demandingness objection, despite having been developed to accommodate agents' pursuit of their own interests, could be extrapolated to accommodate agents' abnegation of their own interests. It is difficult to see how such an extrapolation could be achieved though.

One possibility is to argue that reductions in overall value due to failures to self-benefit are mitigated by other goods involved in the exercise of choice. So in the case of Enrollment, even though Josephine's helping Kelly enroll in the ethics course may be worse overall from the standpoint of well-being (since Kelly benefits less from the course than Josephine would), her doing so has value insofar as it is an exercise of liberty, autonomy, or the like. And if the value of Josephine's choosing is equal to or greater than the value of the well-being she foregoes by helping Kelly enroll, then appearances notwithstanding, Josephine has satisfied the consequentialist standard of maximization. For no other action ranks better in terms of overall good than her helping Kelly enroll in the ethics course.

This reasoning suffers from two defects, First, the value of choice will, in general, contribute as much value as agents forego when they fail to benefit themselves is an unlikely thesis. Again, suppose in Evacuation that Igor foregoes a great deal of well-being by remaining in his apartment. There is no obvious reason to suppose that this amount of well-being is counterbalanced by some equal or greater quantity of well-being associated with his exercising her power to choose not to benefit himself. Second, recall that the permissibility of failing to benefit oneself rests on an option: It is permissible to (p. 263) benefit oneself or to forego such benefits. This reasoning cannot make sense of such an option. For suppose in Enrollment that Josephine decides instead to enroll in the ethics course herself. In order for this to be a permissible option on a consequentialist view, it must (with respect to overall good) tie with her actual decision to help Kelly enroll in the course. But that appears impossible. For assuming that whatever contribution her choosing makes to the overall value of the ensuing state of affairs remains steady (i.e., that value is the same regardless of which option she chooses), then because her own enrollment promotes the greatest amount of well-being overall, then her enrolling must produce the greatest good overall. The only way to avoid this conclusion is to prove the unlikely claim that when Josephine helps Kelly enroll, the value of her so choosing exceeds the value associated with her choosing to enroll herself.

The other general route for addressing the demandingness objection is accommodationist—to establish an "agent-centered prerogative" that allows agents not to "devote energy and attention to their projects and commitments *in strict proportion* to the value from an impersonal standpoint of their doing so."[16] As developed by Samuel Scheffler, the agent-centered prerogative is intended to enable agents to permissibly pursue their central

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

projects and commitments even when doing so would, from a strictly impersonal point of view, not be maximally good.

The agent-centered prerogative thus seems to license morally permissible failures of self-benefit, since they too are deviations from what would be maximizing from an impersonal point of view.

Here again, that failures to self-benefit are exercises of moral options stymies consequentialist efforts to answer the compulsory self-benefit objection. If the appeal to an agent-centered prerogative amounts to asserting that when agents fail to benefit themselves they do not maximize overall well-being but do realize other goods—the value of choice, autonomy, integrity, and so on—then this is simply a restatement of the strategy we just rejected. But if it is not an appeal to the value of choice, it is hard to see that the considerations that motivate an agent-centered prerogative allowing individuals to forego maximizing good in the service of their central projects and commitments support the moral permissibility of an option not to benefit oneself. For the intuitive basis of this option is not a moral permission *not to maximize* because of some compelling reason that emerges from within individuals' personal points of view. Consider Evacuation again: Igor presumably does not maximize goodness by remaining in his apartment. But the moral permissibility of his doing so is not faithfully captured by the thought he thereby permissibly fails to maximize overall well-being. Scheffler's agent-centered prerogative allows agents not to maximize impersonal goodness by taking into account agent-centered reasons rooted in their personal projects and commitments. The picture suggested here is that when such reasons are arrayed against impersonal reasons, they will at least sometimes (but not necessarily) be sufficiently compelling to establish a moral permission for agents not to maximize impersonal goodness. Examples such as Evacuation remind us that the moral permission not to benefit (p. 264) ourselves is very wide, not an option to *discount* our well-being to some degree but an option to exempt our well-being, partially or in full, from the moral calculus as we see fit. Igor's permission not to benefit himself does not flow from any judgment regarding whether he correctly balances impersonal reasons with agent-centered ones. It instead rests on a *right to disregard* his well-being for moral purposes, an entitlement to set aside his well-being so far as moral decision making goes. We prescind from moral criticism of choices like Igor's from a recognition that his relationship to his good is his business. Agents thus enjoy a sort of *authority* with respect to their own well-being. This authority may be cashed out in terms of what Joseph Raz called "exclusionary reasons," reasons "to refrain from acting for some reason."[17] When a commanding officer gives a soldier a binding order to X, the order serves as a reason that "excludes" whatever reasons the soldier might otherwise have that bear on X, mooting those reasons in the soldier's deliberation. In like manner, I suggest, the moral permissibility of not benefitting ourselves flows from an authoritative relation we have to our well-being, one that permits us to exclude our well-being from the domain of moral appraisal—to treat our own well-being as irrelevant to moral choice. Note that the permissibility of failures to self-benefit is itself a *moral* permission; for it would be morally objectionable in most cases to compel individuals to benefit themselves. But it is a permission not rooted in the quality or magnitude of a person's reasons, impersonal or agent-centered, but in a

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

basic moral *power* to exclude one's own good from the practical determination of what is morally best or obligatory.[18]

Consequentialist responses to the demandingness objection are therefore unlikely to succeed in addressing the compulsory self-benefit objection: if they appeal to how failures to maximize (say) overall well-being can be counteracted by other goods, then it is unlikely that these other goods are just valuable enough to establish an option between benefitting ourselves and failing to do so. And effective ties between these options are unlikely given the apparently wide breadth of permissible failures to self-benefit. If consequentialists attempt to extend the agent-centered prerogative to failure to self-benefit, this incorrectly grounds the permission not to benefit ourselves in the balance of reasons among impersonal and agent-relative reasons. The evidently wide permissibility of not benefitting ourselves, I propose, appears to instead be rooted in a moral power to exclude our own well-being from moral deliberation and choice.

5. Dual-Ranking Consequentialism

A final theoretical option for addressing the compulsory self-benefit objection is dual-ranking act consequentialism. The theory and its philosophical motivations are too (p. 265) complex to investigate in depth here. But the gist of the theory (as articulated by Douglas Portmore[19]) is as follows: Because they morally evaluate actions in terms of their outcomes, all consequentialist theories are necessarily committed to ranking outcomes in terms of their being better or worse. Orthodox versions of consequentialism rank outcomes in terms of value or goodness from an evaluator-neutral point of view. Portmore's dual-ranking theory diverges from these versions of consequentialism in two ways. First, outcomes are ranked not in terms of the goodness or value resulting from an action but in terms of the desirability of outcomes. Second, outcomes are ranked both in terms of moral reasons (i.e., reasons rooted in what outcomes would be better for others[20]) and nonmoral reasons, with the latter including what are standardly thought of as agent-relative reasons, such as a person's reason to want to not cause harm to others (understood as distinct from the agent-neutral reason not to want harms to occur). By incorporating these nonmoral, personal reasons, dual-ranking theory provides a ranking that is relative to particular evaluators or agents. Moral permissibility, on Portmore's picture, does not turn solely on moral reasons. For given the truth of moral rationalism—that agents can only be morally required to do what they have decisive reasons to do, all things considered—it may be the case that agents have sufficient reason not to do what moral reasons alone mandate. Moral permissibility thus turns on both moral reasons and an agent's all-things-considered reasons, so that an act is morally permissible for a given agent "if and only if, and because, there is no available act alternative that would produce an outcome that [the agent] has both more moral reason and more reason, all things considered, to want to obtain."[21] Dual-ranking act consequentialism appears capable of answering the compulsory self-benefit objection because it provides agents morally permissible options when outcomes diverge with respect to moral versus all-things-considered reasons. The option not to benefit oneself arises when an agent has a nonmoral reason to benefit her-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

self[22] such that this reason, in concert with her other reasons, entails that she has most reason all-things-considered to benefit herself but most moral reason not to benefit herself. Thus, in examples such as Enrollment, we may view Josephine as (a) having more moral reason to enable Kelly's enrollment, since that results in the better outcome for others, but (b) most reason all-things-considered to enroll herself. By making logical space for agents to act on options that do not maximize goodness from an all-things-considered perspective, dual-ranking consequentialism looks especially promising in addressing the compulsory self-benefit objection. That said, this strategy faces difficulties on two fronts.

First, dual-ranking act consequentialism analyzes options in terms of divergences between moral and nonmoral (or between moral and all-things-considered) agents' reasons. But we may wonder whether all instances of failure to self-benefit can be analyzed in this way. In Enrollment, Josephine may well face a situation in which her moral (p. 266) reasons point one way and her nonmoral reasons another way. Perhaps, then, dual-ranking act consequentialism fares well in accounting for cases of nonoffsetting failures of self-benefit. But it appears shakier with respect to cases of pure failures of self-benefit, such as Igor in Evacuation. Again, we may be curious as to what Igor's reasons for not evacuating and so causing himself harm are. But his remaining being permissible does not seem to be a matter of his having *more* reason all-things-considered to remain in his apartment, reasons in comparison to which his moral reasons are comparatively modest. Only *his* good is at stake. It looks as if his moral and nonmoral reasons align here such that the permissibility of his not benefitting himself cannot be traced to any facts about how weighty those reasons are in relation to one another.

More generally, dual-ranking act consequentialism, even when it logically implies permissible failures of self-benefit, may not provide the most parsimonious explanation of the option not to benefit oneself. Portmore dubs his dual-ranking consequentialism "common sense" inasmuch as it recognizes that moral reasons are not rationally decisive. But I doubt that "common-sense" reactions to cases of failure to self-benefit would judge them permissible *because* in such instances, an agent has no other act alternative available to her that "would produce an outcome that [the agent] has both more moral reason and more reason, all things considered, to want to obtain." As we noted in the previous section, Igor's failure to benefit himself is immune to moral criticism, most would say, because his not benefitting himself is his right, an option to which he is entitled because he is deciding about his own good instead of the good of others. Other moral agents do not so much judge his act as morally permissible in light of his reasons as they do prescind from judging his reasons at all. For like other competent moral agents, Igor's relationship to his good (and Josephine's to hers) is largely his business, and while he may sometimes be entitled to prioritize his good, he is no less entitled to deprioritize his good without reference to the first-order reasons that motivate his deprioritizing it.

Dual-ranking act consequentialism errs, I suggest, in trying to account for options such as the permissibility not to benefit oneself by reference to agents' first-order reasons for action. It is probably correct to deny that moral *reasons* necessarily give agents decisive

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

reasons for action and so exhaust the factors that determine acts' deontic status. But the permissibility of not acting on what there is most moral reason to do, including failing to benefit oneself, is explained more directly, simply, and elegantly in terms of our having a moral power or authority over ourselves rather than in terms of conflicts between two categories of reasons and the relative magnitudes of the reasons within those categories.[23]

In fairness, dual-ranking act consequentialism could incorporate the power to exclude one's own good from moral consideration by thinking of this as a second-order reason. Some of our reasons, after all, are reasons rooted in such powers (some moral philosophers would classify these as "reasons of autonomy," etc.). But introducing (p. 267) second-order reasons threatens to complicate an already complex account of moral permissibility. For the theory must then explain how first-order moral reasons, first-order non-moral reasons, *and* second-order moral reasons (which in turn shape the role first-order reasons play in determinations both of an agent's all-things-considered reasons and of moral permissibility) relate in such manner as to yield options when moral reasons and all-things-considered reasons diverge. It would be premature to claim that such relations cannot be plausibly elucidated, but some a priori skepticism about that project seems warranted.

6. Conclusion: Self, Other, and Directed Options

Our discussion has canvassed some, but not all, of the possible consequentialist responses to the compulsory self-benefit objection that attempt to establish the permissibility of not benefitting ourselves. While these responses vary in their shortcomings, their struggles in addressing this objection help illuminate why the objection is troubling for consequentialism.

At its heart, consequentialism stands opposed to actions having fundamentally *directed* deontic status.[24] As debates about special obligations indicate, in claiming that our duties rest on bringing about particular outcomes, consequentialism struggles to account for how the performance of our duties can be owed to specific individuals or how failures of duty wrong them. After all, having a duty to a person is crucially different from having a duty to realize some state of affairs. The compulsory self-benefit objection shows that consequentialists similarly struggle to make sense of directed *options*: the permission not to benefit oneself is an option but not one that an agent has with respect to anyone beside herself; that is, she is not at liberty to assign others' good lesser significance in her moral deliberation. A consequentialist response to the compulsory self-benefit objection would therefore need to invoke some sort of asymmetry between oneself and others to make sense of it as an option.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Moreover, as sections 4 and 5 illustrate, the wide breadth of the moral permissibility of not benefitting ourselves implies that consequentialist approaches that try to answer this objection by appealing to the strength of personal (or nonmoral) reasons misrepresent the nature of this permissibility. It rests not on some category of (first-order) reasons whose significance permits us not to benefit ourselves but on a seemingly more basic moral power or right to exclude, to whatever degree an individual sees fit, her good from the deliberative weighing of reasons.

(p. 268) In sum, then, the compulsory self-benefit objection resists an easy consequentialist answer because it requires much more than simply making sense of nonmaximizing options. In resting on a directed option, it exerts pressure on consequentialists' commitment to impersonality, and in having a wide breadth, it exerts pressure on the fundamental consequentialist assumption that all and only outcomes of actions contribute to their deontic status.

Notes:

([1]) Stephen Finlay, "Too Much Morality?" in *Morality and Self-Interest*, edited by Paul Bloomfield (Oxford: Oxford University Press, 2008), 142.

([2]) Peter Singer, "Famine, Affluence, and Morality," *Philosophy and Public Affairs* 1 (1972): 229–243.

([3]) James Rachels, "Morality, Parents, and Children," in his *Can Ethics Provide Answers?: And Other Essays in Moral Philosophy* (Lanham, MD: Rowman and Littlefield, 1996), 213–234.

([4]) John Harris, "The Survival Lottery," *Philosophy* 50 (1975): 81–87.

([5]) Michael Slote, "Some Advantages of Virtue Ethics," *Identity, Character, and Morality*, edited by Owen Flanagan (Cambridge, MA: MIT Press, 1990), 441.

([6]) See Peter Singer, *Practical Ethics*, 3rd ed. (New York: Cambridge University Press, 2011), 20–24, for a canonical expression of this principle of "equal consideration of interests."

([7]) I engage with some strategies not addressed here in my "Agents, Patients, and Compulsory Self-benefit," *Journal of Moral Philosophy* 11 (2014): 159–184.

([8]) See Robert Merrihew Adams, "Motive Utilitarianism," *Journal of Philosophy* 73 (1976): 467–481.

([9]) Diana T. Meyers, "The Politics of Self-Respect: A Feminist Perspective," *Hypatia* 1 (1986): 83–100.

([10]) Ted Sider, "Asymmetry and Self-sacrifice," *Philosophical Studies* 70 (1993): 117–132.

([11]) Sider, "Asymmetry and Self-sacrifice," 128.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([12]) I take such reasoning to also speak against "sophisticated" consequentialist attempts to answer the compulsory self-benefit objection. See my "Agents, Patients, and Compulsory Self-benefit," for more discussion.

([13]) See, in a very large literature, Samuel Scheffler, *The Rejection of Consequentialism* (Oxford: Clarendon/Oxford, 1984); Liam B. Murphy, "The Demands of Beneficence," *Philosophy and Public Affairs* 22 (1993): 267–292; and Tim Mulgan, *The Demands of Consequentialism* (Oxford: Oxford University Press, 2001).

([14]) David Sobel, "The Impotence of the Demandingness Objection," *Philosophers' Imprint* 7 (2007) www.philosophersimprint.org/007008/, 2.

([15]) Samuel Scheffler, *Human Morality* (Oxford: Oxford University Press, 1992), 98.

([16]) Scheffler, *The Rejection of Consequentialism*, 9–10.

([17]) Joseph Raz, *Practical Reason and Norms*, 2nd ed. (Princeton, NJ: Princeton University Press, 1990), 39.

([18]) For an elaboration of moral agency in terms of practical powers, see Michael Cholbi, "Paternalism and Our Rational Powers," *Mind* 126 (2017):123–153.

([19]) Douglas Portmore, *Commonsense Consequentialism: Wherein Morality Meets Rationality* (Oxford: Oxford University Press, 2011).

([20]) Portmore, *Commonsense Consequentialism*, 94.

([21]) Portmore, *Commonsense Consequentialism*, 118.

([22]) Portmore, *Commonsense Consequentialism*, 40.

([23]) See my "Agents, Patients, and Compulsory Self-benefit," section VII, for further details about the powers in question.

([24]) See Marcus Hedahl, "The Significance of a Duty's Direction," *Journal of Ethics and Social Philosophy* 7, no. 3: 1–29, for discussion of deontic directedness.

Michael Cholbi

Michael Cholbi is Professor of Philosophy at the University of Edinburgh. He has published widely in ethical theory, practical ethics, and the philosophy of death and dying. His books include Suicide: The Philosophical Dimensions (Broadview, 2011), Understanding Kant's Ethics (Cambridge University Press, 2016), and Grief: A Philosophical Guide (Princeton University Press, expected 2021). He is the editor of several scholarly collections, including Immortality and the Philosophy of Death (Rowman and Littlefield, 2015), Procreation, Parenthood, and Educational Rights (Routledge, 2017), The Future of Work, Technology, and Basic Income (Routledge, 2019), and The Movement for Black Lives: Philosophical Perspectives (Oxford University Press,

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

2020). He is the founder of the International Association for the Philosophy of Death and Dying and the coeditor of the textbook Exploring the Philosophy of Death and Dying: Classic and Contemporary Perspectives (Routledge, 2020). His current research addresses paternalism, assisted dying, and topics related to work and labor.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Consequentialism, Blame, and Moral Responsibility

Elinor Mason

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.19

Abstract and Keywords

In this chapter I examine various accounts of the relationship between consequentialism and moral responsibility. The first idea is that the only reason we have for praising and blaming, for holding responsible, is that it will produce good consequences. This view is widely derided, but a descendant, the view that our responsibility practices as a whole can be defended on consequentialist grounds, has been gaining popularity in recent years. I go on to look at the idea of blameless wrongdoing and give an account of how that might fit into to a consequentialist picture. Finally, I discuss the possibility that the direction of influence is the other way: that consequentialist ethical theories are constrained by theories of moral responsibility, and I discuss possible upshots of a responsibility constrained account of consequentialism.

Keywords: blame, blameworthiness, praise, praiseworthiness, responsibility, blameless wrongdoing, objective rightness, subjective rightness, prospectivism

THEORIES of moral responsibility and theories of ethical principles are surely connected in some way. It is not entirely clear how they are connected though. Do consequentialists about ethics have reasons to accept a consequentialist account of moral responsibility? Or are the two things entirely independent? Or, conversely, do accounts of moral responsibility limit what consequentialists can say about right action? In this chapter I survey the various possible views, in the hope of illuminating the relationship between consequentialism about ethical principles and theories of responsibility and blame.

1. Consequentialist Accounts of Moral Responsibility

A consequentialist account of morality says that actions are right or wrong as they have good or bad consequences. This view is much discussed, and much maligned, but also frequently defended. A consequentialist account of morality seems to be respectable at least

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

—not a nonstarter anyway. A consequentialist account of moral responsibility, by contrast, seems doomed. But perhaps reports of its death have been exaggerated.

The first generation of consequentialist accounts of moral responsibility said that people are responsible—praise or blameworthy—when holding them responsible would have good consequences. The view appears in Sidgwick (1907, 71–72), but it is mainly associated with later theorists. Both Moritz Schlick (1939) and J. J. C. Smart (1961) argue in a brisk way for the view that the *only* reason we might have for praising and blaming is the good consequences of doing so. In particular, we praise and blame to (p. 163) influence people's behavior in a desirable way. They might be right that insofar as praising and blaming are *actions*, consequentialists should take the value of the consequences of performing those acts as the relevant factor in deciding whether or not to perform them.[1] But in the more important sense of praise and blame (or so it seems), praise and blame are not acts but at least partly *attitudes* and, thus, are like beliefs, in that giving consequentialist reasons for them seems mistaken. It appears to be essential to our practices of praising and blaming that we look *backward*, not forward, that we can point to something the agent did that justifies our reaction, not to something that she will do in the future.

In sum, the consequentialist account of responsibility is extremely counterintuitive. We have a very strong pretheoretical conviction that we should hold people responsible when they *are* responsible, and any other justification gets things badly wrong. A consequentialist justification for holding responsible strikes us as inappropriate in the same way that a consequentialist justification for beliefs is inappropriate. We face a classic "wrong sort of reasons" problem. The claim that responsibility can be understood in a consequentialist way seems like a nonstarter.[2]

However, it is worth pausing to think about the motivation for the consequentialist account of moral responsibility proposed by Schlick and Smart. Moral responsibility presents a puzzle, and the "moral influence account," as it is often known, solves that puzzle. The puzzle, of course, is the puzzle of free will and determinism. The obvious answer to the question "when should we hold someone responsible?" is "when they acted freely." But if determinism is true, it is never the case that people act freely in the sense pretheoretically intended here. So we need another story about when we should hold people responsible, if at all. The influence story is compatible with determinism. We should praise and blame people when that will affect their behavior in positive ways. Further, Smart says that to be a moral agent is to be influenceable in this way. We cannot affect the behavior of fence posts, but we can affect the behavior of agents. The story is neat and metaphysically unimpeachable. Sometimes we have to give up on our ambitious pretheoretical ideas and accept a deflated version. If determinism is true, this must be the case with moral responsibility.

Nonetheless, the influence account seems problematic. It is not the right sort of story, and it takes us too far from our pretheoretical ideas. What we want, when we run into a problem like determinism, is to preserve as much as we can of the valuable practices that are threatened. The Schlick/Smart view is overly revisionist. Our practices as they are de-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

pend on us thinking that we hold people responsible when they are responsible in some sense that is independent of the consequences of holding them responsible.

If there were no compatibilist account that could make sense of the backward-looking nature of moral responsibility, then we might have to be satisfied with the influence (p. 164) view. But P. F. Strawson's influential suggestion, that we should look at the agent's quality of will in acting, provides an alternative (1962). As Strawson points out, that someone acted in a mean or vicious way is what we blame him for.[3]

As Strawson puts it:

> If someone treads on my hand accidentally, while trying to help me, the pain may be no less acute than if he treads on it in contemptuous disregard of my existence or with a malevolent wish to injure me. But I shall generally feel in the second case a kind and degree of resentment that I shall not feel in the first. If someone's actions help me to some benefit I desire, then I am benefited in any case; but if he intended them so to benefit me because of his general goodwill towards me, I shall reasonably feel a gratitude which I should not feel at all if the benefit was an incidental consequence, unintended or even regretted by him, of some plan of action with a different aim. (1962/2003, 76)

Strawson argues that this is how our practice of responsibility actually works and then points out that the truth of determinism would not undermine this practice. That the person was determined in this way does not undermine the difference between an intentional injury and an accidental injury.[4] Whether or not we think that this is *all* there is to say about responsibility, this is certainly something we *should* say about responsibility. If some effect occurs through an agent's body (she hurts me or helps me) that is not connected with her agency (she is blown by the wind), we would not think ill or well of her. If, on the other hand, she intended to help me, and moved her own body so as to achieve that, our reaction is, appropriately, a reaction to the agent *qua* agent, and, as Strawson says, our reactive attitudes in these cases are the attitudes that constitute our responsibility practice, praise and blame.

Should we worry that the action is not free? Here we can look to Frankfurt, the co-parent of contemporary compatibilism (1969). As Frankfurt points out, we think of some acts as being the agent's own, in that it was not some other agent interfering, but just the agent's own motives that led to the action. Frankfurt gives examples that are designed to convince us that we already think that it doesn't really matter if an agent could have acted differently, so long as she identifies with her action and takes it to be her own. Between them, Strawson and Frankfurt provide a foundation to build a nonconsequentialist compatibilism, one that gives us the backward-looking element of (p. 165) moral responsibility without the metaphysical ambition. Subsequent to Strawson and Frankfurt's interventions, a wide variety of compatibilist accounts of moral responsibility has developed along these broad lines.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

This still leaves a question about why we would prefer preserving the practice to abandoning it (assuming, perhaps wrongly, that we could abandon our practices of holding people responsible). Strawson and Frankfurt point the way to a version of our responsibility practice that is internally consistent and does not rely on a libertarian view about free will. However, it is not clear that we have any reason to accept the new version of the practice.[5] This is where consequentialism comes back in.

R. B. Brandt was perhaps the earliest critic of the influence account to suggest a consequentialist alternative (1969). Brandt feels the force of the standard objections but proposes a rule utilitarian account of holding people responsible. On Brandt's view, a crucial aspect of the rule utilitarian view is that we need to think about the benefits of a moral code being internalized. The moral code includes a story about what counts as an excuse —about when an otherwise wrong action should not be condemned. So the rule utilitarian should think about the best (most utility producing) account of when an agent is excused. This, of course, is just the other side of the praise and blame coin: Brandt's question is "when we should blame people?" Brandt argues that we need to think about gains and losses—if we allow too few excuses, the code will be hard to internalize and there will be costs in terms of guilt feelings and self-respect. On the other hand, with too many excuses the moral code will be insufficiently demanding and will not have good effects on behavior.

Brandt suggests the moral system (unlike the legal system) does not work primarily by threats, but by internalization, by aversions. That has ramifications for what system of excuses (blame) is going to be most effective:

> There is no point, in general, in a moral system condemning failure to do the impossible, or the accidental, or what is done when a person is hypnotized or paralyzed by fright. If the utilitarian theory of excuses is correct, such considerations must entirely exculpate from blame. The preventive capacities of the moral machinery are in no way reduced by excepting such kinds of behavior from moral blame. (1969, 353)

Brandt ends up with an account that is broadly similar to Strawson's in the details of when we should hold people responsible. He says that the system of excuses should excuse when the act does not manifest a defect of character, or defect in motivation, where there is some ideal level of motivation that can be defined by thinking about what level of motivation has the highest utility. Brandt also says that being subject to the guilt and disapproval of others increases motivation in the desired direction and strengthens that trait of character (357). This thought, that the moral responsibility (p. 166) system works to make people better as moral agents, features heavily in more recent consequentialist accounts.

Recent theorists (including Arneson 2003; Jefferson 2018; McGeer 2015; Vargas 2013) argue for a broadly consequentialist justification for responsibility practices, where what is being justified is not praise or blame of particular acts, but the responsibility system as a whole, where that system may well include nonconsequentialist elements, and even ves-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

tiges of a commitment to free will. Manuel Vargas presents a thoroughly worked-out account of the practices that are justified in consequentialist terms, including an account of moral agency, development of which, Vargas argues is one of the consequences that the practice justifies.[6]

This two-level structure raises an issue of interest to consequentialists: just as critics of consequentialism are not mollified by the move to indirect consequentialism, unimpressed by the argument that it is not normally permissible to frame the innocent on the grounds that a practice of punishing only the guilty has better consequences, so a critic of the moral influence view of praise and blame might remain unimpressed by the move to justification at the level of practice. The worry is that the consequentialist justification pollutes; it leaks through the justification levels and renders the permission to proceed in an apparently nonconsequentialist way hollow. It seems to miss the point to say that the reason not to frame an innocent person is that the practice has the best consequences. The consequentialist's opponent insists that the reason not to frame the innocent is simply that they are innocent. The complaint is that that pretheoretical idea seems worth preserving in our theorizing, and the consequentialist does not succeed.

In the case of moral responsibility practices, the pretheoretical idea that we wanted to preserve was that we should hold someone responsible when she *is* responsible. Second-generation consequentialist accounts of moral responsibility seem no closer to securing that than the moral influence view. And we are back to square one: if we want to adhere to a naturalistic worldview, we have to substantially modify our ambitions for our responsibility practices. On the new improved consequentialist account, our responsibility practices may *look* much as they did, but if we accept the consequentialist story about justification for our practices, we are a long way from the presuppositions of the original practice.[7] Thus the same worry might arise, that despite the appearance of preserving what was pretheoretically attractive, a consequentialist justification structure undermines any claim to preservation.

(p. 167) However, the proponent of the new improved consequentialist account need not be perturbed by this. The presuppositions of the original practice, were, after all, false—there is no metaphysical free will. So dialectically, we need to set that aside at the beginning. The question is not, "is this view superior to a libertarian account of our responsibility practices?" The question is, rather, a question in normative responsibility theory: "can we come up with a sensible, internally coherent account of something close to our practices (close enough not to be changing the subject), that does not advert to metaphysical free will?" Arguably, the new improved consequentialist view does a good job. The question then is how it compares to other compatibilist accounts on the desiderata like coherence and consistency. We should not worry that the new improved consequentialist story about our responsibility practices is revisionist. There is no sense to the pretheoretical version of the thought that the agent should be held responsible because she *is* responsible.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

These observations about the dialectic contrast in interesting ways with the analogous argument in consequentialist ethical theory. In both ethics and moral responsibility theory we can set aside questions about realism and anti-realism. We can say, if we wish, right from the start, that there is nothing metaphysical on the table; all our values and norms are part of the natural world. This does not lead straight to nihilism. We can tell stories about why our practices make sense. These naturalistic stories will of course be limited in some ways. Our moral- and responsibility-related ideas like "goodness" and "badness," "evil," "blameworthiness," "desert," and so on will have to be given slightly revisionist readings—pruned of any taint of nonexistent referents. So "good" doesn't refer to any nonnatural property, but just to some descriptive property in the world, the happiness of sentient beings, perhaps. "Desert" doesn't refer to a magical property that does justifying work; rather, it summarizes the relevant descriptive facts. If we decide that we ought to punish someone when they acted in full knowledge of the badness of their act, then that is all that it is to deserve punishment.

So far so good. The move to indirect theory in talking about moral responsibility—to justifying the practice rather than the individual acts of praise and blame—does a lot of work. It brings the practice closer into line with what seems intuitively to make sense. And the criticism "but the justification in the background pollutes" does not apply, because there is no alternative. The so-called polluting is just the absence of metaphysical free will.

However, the picture in the case of the move to indirect justification in ethical theory is rather different. There, the claim that we started with, the pretheoretical intuition, is (to stick to the same example) that we should not frame the innocent, just because they are innocent. In this case, we do not have to abandon the original thought; it is not the case that there is no such thing as innocence. Of course, we might like an account of why innocence matters, and there is a question about how far that story has to go. But there is no principled reason not to say that innocence (in the sense of not having done the crime in question) is bedrock, that one of the axioms is "only punish the person who did the crime." So we can see why the worry that consequentialist justifications bleed between levels has more bite in the ethical case. Punishing people when it would be (p. 168) expedient to seems misguided, but arguing that a practice of punishing only the guilty is expedient seems equally misguided.[8]

It is interesting, then, that indirection (the move to a higher level of justification) is successful in consequentialism about moral responsibility, and not always so much in ethical theory. Are there any general lessons for ethical consequentialism here? I think it is interesting to note that there are variations in the sorts of pretheoretical ideas that feed into the process of reflective equilibrium. Some are candidates for being bedrock, some are not, and some can *only* be accommodated in an indirect way. If our process of moral theorizing is sensitive to our pretheoretical intuitions, then we can't always solve problems of counter-intuitiveness by arguing for an indirect approach.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

2. Blameless Wrongdoing

I will very briefly comment on the phenomenon of "blameless wrongdoing." The term was introduced by Parfit (1984), and it applies to cases where an agent has an option that includes, as an essential part, an action that, when compared atomistically to other actions that the agent could be doing as part of an overall less good option, is less good. In Parfit's example, Clare has a choice between benefiting her child and benefiting a stranger, and she decides to benefit her child. She does this, according to Parfit, because she has the best possible set of motives (i.e., the set of motives which results in the best consequences). However, Parfit claims that in causing Clare to benefit her child rather than the stranger, these motives cause her to do something that will make the outcome worse. Parfit wants to acknowledge that there is a sense in which this action is suboptimal, but also maintain that it should be done, so he comes up with the phrase "blameless wrongdoing."[9] Understood in this way, it is not clear that "blameless wrongdoing" really has anything to do with blame or blameworthiness.

There is, of course, an issue about moral responsibility lurking here, which is that sometimes we cannot separate out the parts of a course of action, and so we cannot do the action as considered atomistically: perhaps Clare could not both have good motives and benefit the stranger. That is not what Parfit has in mind though. He imagines that Clare *could* benefit the stranger, but that it would be very hard for her, given her good maternal motives. Parfit says that she is acting wrongly in benefiting her child "only in a very weak sense" (1984, 33).

This raises a general issue about the relationship between difficulty and excuses: all theories of responsibility and right action must give an account of what qualifies as an excuse, and whether or how difficulty is a mitigating circumstance. But there is another way to see the point here, which applies particularly to consequentialists. There might (p. 169) be a special reason to avoid blaming certain kinds of wrong action. This may or may not bring us back to something like the moral influence account of praise and blame.

Katarzyna de Lazari-Radek and Peter Singer (2014) make a suggestion along these lines in their discussion of Sidgwick's view, that there are two different questions to ask: "what a man ought to do or forbear" and "what other men ought to blame him for not doing or forbearing" (1907, 221). Lazari-Radek and Singer are looking for a response to the demandingness problem. The idea is that perhaps we can say that sometimes we act wrongly in not doing more to meet the demands of consequentialism, but we should not be blamed.

Sidgwick, as has already been noted, tends to the moral influence view, that is, that all that there is to justify praise and blame are the effects. But we might of course think that there are some cases where the results of praising and blaming should be taken into account without that generalizing to the conclusion that the moral influence theory of praise and blame says all that there is to say. Lazari-Radek and Singer suggest that Parfit's Clare is one such case (2014, 332). We should not blame her, even though she

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

could have helped the stranger, because we should encourage her to have the motives that she does. The choice she is faced with is, presumably, an unusual and exceptional situation. Normally, we can rely on our good parental motives to have good results, but in this case, a gap has opened up between the general tendency of a particular motive set and the effects on a particular occasion. One way to address this is to say that although usually we should praise or blame along nonconsequentialist lines, for example, based on whether someone acts wrongly knowingly or through bad will, there are exceptions. Clare may be acting wrongly knowingly (assuming she knows that she could benefit the stranger), but nonetheless, the consequentialist might argue that she should not be blamed, and not just because it would have been difficult for her to act rightly. In general, it is good for Clare to have strong parental motives, and so she should not be blamed for acting on them.[10]

This does not commit us to saying more generally that all that justifies praise and blame are the consequences of praising and blaming. We might be committed to a nonconsequentialist practice of responsibility on consequentialist grounds (as Vargas argues) and yet allow that there are triggers that switch us back to a consequentialist rationale for praise and blame. The cases where praise and blame should switch to consequentialist can be picked out as exceptional: as when the usual effects of a disposition diverge from the actual effects. This gives us a mixed theory, where blame sometimes but does not always follow culpable wrongdoing.

I have briefly surveyed the ways in which moral responsibility theory might be consequentialist. I now turn to a less developed area, the question of how the ethical theory of consequentialism may be constrained by moral responsibility theory.

(p. 170) 3. Moral Responsibility Constrained Accounts of Consequentialism

In its simplest form, consequentialism says just that the right act is the one that maximizes utility. This is a feature of the act that may not be accessible to the agent, and so, by any reasonable account of moral responsibility, not one the agent is likely to be responsible for doing. However, those who defend this view maintain that the right act is the one with the best consequences, and that is true and important regardless of whether the agent can do that act, or can do it intentionally, or whatever else might be thought relevant to responsibility. The agent may well have an excuse for not doing the act (she couldn’t have known which one it is), but that doesn’t affect the consequentialist account of right action.

The question is whether we should think that our moral concepts are tied to our responsibility concepts. Kant implicitly argues that they are, and his view is compelling. As Kant famously says in *The Groundwork* (1785), the only thing that is good without qualification is the good will.[11] Kant’s examples of people whose acts have good features that were not strictly intended by them illustrate his point clearly: for Kant, intention is the fundamen-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

tal measure of whether an agent has acted rightly. There is no room for luck. An act is only right if the agent is praiseworthy for it, and conversely, she is always blameworthy for acting wrongly, because to act wrongly just is to act from a bad will.

In terms of the connection to moral responsibility, the traditional consequentialist account of right action is at the opposite end of the spectrum from Kant's view. Rightness is radically divorced from praiseworthiness. The right action is the one with the best consequences, and that action may or may not be accessible to the agent. The best consequences may be entirely unpredictable from the agent's perfectly reasonable point of view. An extreme proponent of this version of consequentialism may have to concede that we cannot blame an agent who acts wrongly in circumstances when he is doing his sincere best.[12] So consequentialists, like any other ethical theorist, must take a position on how much is the concept of rightness constrained by or, alternatively, separable from our account of praise and blameworthiness.

In recent work I put this in terms of a "responsibility constraint" (Mason 2019). There is a responsibility constraint on rightness, such that there must be some connection between right and wrong action and responsibility, or praise- and blameworthiness. We obviously all agree that there is *some* connection (we don't call things right unless they are the sort of things we could be responsible for). There is a huge difference between mere grading (which is how Smart's view is often described) and *moral* appraisal. (p. 171) Whether an avalanche has good consequences or bad consequences is all that matters in our appraisal of an avalanche because an avalanche is not an agent—not the sort of thing that could have the relevant control or quality of will for a truly moral appraisal. Agents, however, do have the properties that are relevant (even to a compatibilist account of responsibility of course), and so to treat them as we treat avalanches, to think of right and wrong action as being merely a matter of the value of the consequences of what they do, is to ignore the crucial issue.

Recent debates about whether rightness is "objective" or "subjective" can be understood as debates about how closely connected rightness and responsibility, or praise- and blameworthiness, should be. At one extreme there is the objective consequentialist, who sticks to the claim that the right action is the one that has the best consequences. The rationale for objective consequentialism is roughly this: the goodness of the consequences is the guiding light of consequentialism, the most basic claim about the nature of moral value, and so anything else is secondary, for example, a decision procedure that we should use in the light of uncertainty about what we *really* should do, or an account of when we should praise and blame agents that is something other than an account of right action. At the other end of the spectrum is subjective consequentialism, which ties rightness to the agent's own point of view. For subjective consequentialism, rightness and responsibility are closely connected: the agent will always (or almost always) be praiseworthy for right action and blameworthy for wrong action.[13]

Arguments for the more subjective accounts of rightness are, implicitly or explicitly, appeals to the need to connect rightness and wrongness with an account of what the agent

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

could be responsible for; what is up to her. So, take for example, Frank Jackson's defense of a prospectivist account of rightness: "the fact that a course of action would have the best results is not in itself a guide to action, for a guide to action must in some appropriate sense be present to the agent's mind. We need, if you like, a story from the inside of an agent to be part of any theory which is properly a theory in ethics" (1991, 466–467). We can take Jackson as invoking a version of the responsibility constraint. Jackson's complaint is that if a theory gives us an account of rightness that is not accessible to us, if we can't tell which of our options are right and wrong, the theory cannot not guide our action. One way to understand the demand for action guidance is as a demand that rightness or wrongness be accessible and, hence, the sorts of things we could properly be responsible for. If we don't know which things are right and wrong, we would not be praise or blameworthy for doing them.[14]

(p. 172) Other writers, such as Fred Feldman and Holly Smith are more explicit that more subjective accounts of obligation are tied to praise and blameworthiness. Smith says that the concept of subjective rightness "should bear appropriate relationships to assessments of whether the agent is blameworthy or praiseworthy for her act" (2010, 73). Feldman says, "An adequate practical level principle must provide a way for the agent to avoid at least certain sorts of blame" (2012, 159).

This is one sense in which consequentialist theories are constrained by concerns about responsibility. The idea is that what we can be responsible for limits on what the consequentialist can say about rightness. We cannot be responsible for unforeseeable consequences, so (the argument goes) such things cannot be included in the scope of right or wrong action; rather, we have to restrict the scope of consequentialism to consequences that can be foreseen. This leads to a more ambitious argument about the limits of consequentialism. There is a general principle at work here: that if we cannot be responsible for something, it can't be something a moral theory labels right or wrong. Now, we can all agree that we are not responsible for unforeseeable consequences. But there are other cases where it is less clear what we are or are not responsible for. Take unintended but foreseen side effects, for example—are we responsible for those? If not, then the same principle would seem to show that those things are not within the scope of rightness and wrongness.

The doctrine of double effect is a nonconsequentialist principle that says that there is a morally relevant distinction between what is directly intended and is merely foreseen. So, for example, the killing of civilians may be morally permissible if it is the foreseen side effect of another permissible action, but it may be impermissible if it is directly intended. The consequentialist objects that the consequences are all that matter, and so this distinction is irrelevant. When the results can be foreseen, they should all be taken into account. The way this argument is usually understood, it is consequentialism that drives an account of responsibility, not the other way round. The consequentialist insists that we are responsible for foreseen but unintended side effects, because these are morally important consequences of our actions. The nonconsequentialist response might be put in terms of

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

duties: you have duties to do and not to do certain things, but that things that merely happen (even as side effects of what you do) are not forbidden.

But already we can see that the nonconsequentialist response here is bound up with an account of responsibility. The underlying claim is that we are not responsible for what we merely foresee, but only for what we directly intend. This raises a question: could the consequentialist simply agree? After all, even objective consequentialists ought to agree that if the consequentialist account of rightness includes more than the agent can be responsible for, the agent's praise and blameworthiness is now a different issue. It would be a case of the tail wagging the dog to say that praise and blameworthiness must follow an objective account of rightness. The consequentialist must acknowledge that the limits of praise and blameworthiness are to some extent independent of substantive moral views.

Even if we accept that, there is no easy answer here. Plausibly, the consequentialist was already talking about responsibility in defending the consequentialist answer to the (p. 173) doctrine of double effect. The consequentialist's thought is that if a consequence is foreseen, that is enough for it to be within the agent's responsibility sphere. The disagreement between consequentialists and nonconsequentialists here cannot be solved by appealing to an independent account of responsibility, because the parties were already talking about responsibility and disagreeing about how *that* works. So the question about the doctrine of double effect boils down to a disagreement about responsibility. Consequentialists and nonconsequentialists may disagree about responsibility in ways that reflect their ethical commitments: thinking that the scope of obligation includes the side effects of one's actions would, of course, go hand in hand with an account of responsibility such that we are as responsible for side effects as for what we directly intend. Thus we cannot make progress by switching the argument from ethics to responsibility theory.

There are, however, other ethical theory puzzles in the vicinity where the consequentialist may have more reason to appeal to an independent account of responsibility. The doing-allowing distinction, for example, says that there is a morally relevant difference between what you actually do and what you "allow." Scheffler (2004) offers an argument for thinking that there are grounds based on considerations about agency that force us to accept something like the doing-allowing distinction, no matter what our prior moral convictions. I have offered an argument along similar lines for thinking that that a whole family of objections to consequentialism can be understood as complaining that consequentialism is committed to an implausible account of responsibility (Mason 2018). Bernard Williams (1973) is the source of many of these worries about consequentialism. Williams objects that consequentialism is an attack on the agent's integrity. Unlike the doctrine of double effect (which most consequentialists are happy to bite the bullet on), these objections have motivated consequentialists to try to alter the theory in various ways. A solution that appeals to an independent account of moral responsibility is therefore worth pursuing.[15]

Williams claims that consequentialism is committed to the doctrine of negative responsibility. This is another way of describing the doing-allowing distinction. As Williams puts it,

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

"From the moral point of view [for the consequentialist], there is no comprehensible difference which consists just in *my bringing about a certain outcome* rather than *someone else's producing it*" (1973, 96). Williams's thought here is that consequentialism says that we are responsible for what happens, or what others do, but that is too much responsibility—we are only responsible for what we do. Williams's examples of George the chemist (who must choose between taking a morally suspect job and refusing it, which would result in someone else taking the job and doing it much more efficiently than George would have) and Jim in the jungle (Jim is in a hostage situation and must choose between killing one person himself and allowing twenty to be killed) illustrate the point vividly. Williams's very powerful thought is that it is not *George's* fault (p. 174) if the other chemist is a zealot; it is not *Jim's* fault if Pedro kills twenty people. Why should this be something that we hold George or Jim responsible for? In other words, on Williams's view, the consequentialist approach violates the responsibility constraint: it tells the agent that the right act is one that the agent could not be properly responsible for.[16]

The question, then, is what an independently plausible account of moral responsibility would say. In contrast with the doctrine of double effect, the contested item is not an action of ours that we can foresee effects of, but an action of *someone else's* that we can foresee. So whereas the consequentialist can refuse to accept the doctrine of double effect without being committed to a very controversial account of responsibility (it doesn't seem hugely controversial to say that you are responsible for foreseen side effects of actions), it is much more tenuous to say that agents are responsible for what others do. Of course, the issue is that in the examples where Williams thinks that consequentialism gives too quick an answer, the agent could prevent the other agent from doing the problematic act, but only by herself doing something that she is morally opposed to—hence Williams's claim that consequentialist prescriptions are an attack on the agent's integrity. Consequentialism does not allow the agent to act on his own moral convictions, but rather makes "him into a channel between the input of everyone's projects, including his own, and an output of optimific decision" (116–117). William's point is not that consequentialism gives the wrong answer, but that it misses complexity in the case.[17]

As a theory of ethics, the consequentialist again may be tempted to bite the bullet here. After all, the outcome is so much better (particularly in Jim's case) if we take into account what Pedro will do. Why don't we take that into account? To ignore it seems to fall into a different problem: it seems willfully blind, even self-absorbed, or self-fetishizing. However, an appeal to an independently plausible account of moral responsibility might silence those concerns. Scheffler (2004) offers an account that aims to do just that.

Scheffler starts from a nonskeptical position about moral responsibility: he accepts that we are sometimes responsible on an individual basis. He goes on to argue that accepting that is enough for us to be committed to something in the region of a doing-allowing distinction. The first point is that to accept that one is responsible is not merely to accept that one can be appraised or graded, but to accept that there are standards that apply to one, and to hold oneself to those standards. That involves a commitment to bring one's own conduct into line with the standards of being a responsible agent. And that means

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

that we take our conduct as special: “More generally, to view oneself as subject to norms of individual responsibility is to draw a normatively relevant distinction between one’s own conduct and all other causal processes. It is, in other words, to see oneself as responsible for regulating the exercise of one’s own agency, as opposed to the exercise of anyone else’s agency or any other causal processes, by reference to those norms” (2004, 221). That point seems undeniable: if we are not skeptical about moral (p. 175) responsibility, we are accepting that we are agents, that we are not just part of the causal chain like any other.

The next step is to show that this basic commitment also involves a commitment to a distinction between primary and secondary manifestations of one’s agency (between doings and allowings, in other words). Scheffler points out that bringing one’s conduct into line with the standard is a doing. It cannot be something that one allows to happen.[18]

So, Scheffler claims, to see oneself as a responsible agent is to see an important difference between primary and secondary exercises of one’s agency: the basic demand that the agent who sees herself as responsible recognizes is a demand to do something (222). Thus accepting the distinction between primary and secondary exercises of agency is a presupposition of seeing oneself as a responsible agent. Put more simply than Scheffler puts it, we might say that the basic idea is that a presupposition of the nonskeptical position is that one must do some things, and not merely allow them.

Scheffler addresses what he takes to be an obvious and powerful objection. It could be argued that although the nonskeptical position requires us to accept that we must conform to the standards of being an agent, and take our own agency as ours to regulate, that might involve taking what one allows to be as important as what one does. In other words the content of the norms of responsibility could be anything.

Scheffler responds by imagining how an agent might be expected to feel in a case like Jim’s where he must choose between doing and allowing harm. Scheffler focuses on how the agent should feel about being a target of resentment if he decides to prevent the greater harm. It seems to Scheffler that it would be “psychologically and humanly absurd” (230) for the agent to expect the group not to resent him for the harm he does. And that, Scheffler thinks, is a sign of a deeper problem, that to expect us to see ourselves and our actions as merely instrumental in the great causal flow toward outcomes is deeply misguided. Here, Scheffler’s language is very much along the lines of Williams’s: the basic worry is one about integrity or alienation. Scheffler’s contribution to Williams’s complaint is the claim that seeing one’s agency as instrumental in this way is at odds with taking oneself to be a responsible agent at all—one must see oneself as having a noninstrumental reason to hold oneself and others to the relevant standards.

Scheffler’s position is a moderate one in the end. He readily admits that we bear *some* responsibility for the opportunities to intervene that present themselves. If an agent is put in a position where she could prevent a disaster and does not, she is at least somewhat responsible. The point is that primary exercises of agency are much weightier. Scheffler does not pretend to have an answer about how we weigh up the different cases of

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

contributions by primary agency and contributions by secondary agency. His point is just that there is such a distinction.

My own argument on this topic takes a different approach (Mason 2018). I focus on the role of the other agent, arguing that there is an important difference between cases where the world is accidentally in a shape such that an agent has a choice between doing (p. 176) and allowing, and the case where another agent deliberately sets things up like that. In a case where another agent is setting things up, that agent, the coercer, is responsible, and his agency has swamped, or ruled out, the agency of the person being asked to do or allow. We can appeal to Frankfurt's account of responsibility to defend this claim (1988). The core thought is that although, in the end, everything is just part of a huge causal chain, including our actions, we can pick out which of our actions we are responsible for by focusing on which of our actions we identify with. It is crucial that an agent thinks of the action as her own (this notion could be cashed out in various ways; I leave it quite vague here). But we cannot think of an action as our own if someone else has dominated the situation. So even a consequentialist can say that in situations of this sort appeal to an independently plausible account of responsibility limits the scope of consequentialist prescriptions.

It is worth noting that the coerced agent can still respond to reasons in the situation, reasons given by a theory of the good, but the fact that her agency has been compromised means that her acts can no longer be assessed as right or wrong. Jim can opt to kill the one to save the nineteen because he thinks that doing so would be better. But, on this account, that is not what *consequentialism* tells him to do. Rather, the reasons are given by the theory of value, the fact that it would be good for nineteen to live rather than die. But Jim's act would not be right or wrong, or one that he could be praised or blamed for in the ordinary way.

Of course, this way of using the responsibility constraint covers much less ground than we might hope for. It leaves consequentialism with no general distinction between doing and allowing; it merely makes space for a special case where other agents dominate. Thus in Jim's case we can say that, as Jim is coerced, and therefore not functioning as an autonomous agent, his actions cannot be appraised as right or wrong. However, the case of George is much less clear. Williams thinks of these cases as relevantly similar, in that both George and Jim seem to have consequentialist reasons to do things that they disapprove of. But whereas appeal to Frankfurt's account of identification can deliver the result that Jim is not functioning as a responsible agent, it is less clear that this applies to George. George is not being dominated by another agent. George has to make a choice about his actions, foreseeing what will happen. It is true that part of what he foresees is what others will do, and that he is not responsible for what they do. Nonetheless, the consequentialist principle, that foreseen consequences are relevant, is not easily displaced by considerations about the responsibility of others in this case.

It is possible that we are simply at the intersection of moral responsibility and ethical theory here, and that thinking about moral responsibility gives us a new angle on these prob-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

lems without giving us any independent purchase on them. We might be able to refine our intuitions about cases like George and Jim by thinking both about responsibility and about ethical principles of action, but in the end neither responsibility theory nor ethical theory can give us firm answers, and we simply have to proceed with the general strategy of balancing plausibility and coherence.

(p. 177) 4. Conclusion

I have discussed various connections between consequentialism as an ethical theory and theories of responsibility and blame. Almost no one holds the simple view, that the only reason for praise and blame is to produce good consequences. But consequentialist accounts of the background justification for our responsibility practice as a whole are very plausible. It may even be that consequentialist elements are appropriate in our practice, as one story about blameless wrongdoing suggests. From the other direction, there is reason to think that consequentialism is limited by accounts of what we can be responsible for. This is widely accepted in at least one context: arguments about how objective or subjective our account of rightness is are appeals to a responsibility constraint on rightness. I have suggested that the strategy of appealing to an independent account of responsibility could be taken further and used to justify a doing-allowing distinction.

References

Arneson, R. 2003. “The Smart Theory of Moral Responsibility and Desert.” In *Desert and Justice*, edited by S. Olsaretti, 233–258. Oxford: Clarendon Press.

Bennett, Jonathan. 2008. “Accountability (II).” In *Free Will and Reactive Attitudes: Perspectives on P.F. Strawson’s “Freedom and Resentment,”* edited by Michael McKenna and Paul Russell, 47–68. Farnham: Ashgate.

Bradley, Ben, and Stocker, Michael. 2005. “‘Doing and Allowing’ and Doing and Allowing.” *Ethics* 115: 799–808.

Brandt, Richard B. 1969. “A Utilitarian Theory of Excuses.” *Philosophical Review* 78, no. 3: 337–361.

Feldman, Fred. 2012. “True and Useful: On the Structure of a Two Level Normative Theory.” *Utilitas* 24, no. 2: 151–171.

Frankfurt, Harry G. 1969. “Alternate Possibilities and Moral Responsibility.” *The Journal of Philosophy* 66, no. 23: 829–839.

Frankfurt, Harry G. 1988. *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press.

Jackson, Frank. 1991. “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection.” *Ethics* 101, no. 3: 461–482.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Jefferson, Anneli. 2018. “Instrumentalism about Moral Responsibility Revisited.” *Philosophical Quarterly* 69, no. 276: 555–573.

Kamm, F. M. 1999. “Responsibility and Collaboration.” *Philosophy and Public Affairs* 28:169–204.

Kant, Immanuel. [1785]. 2002. *Groundwork for the Metaphysics of Morals*. Oxford: Oxford University Press.

Mason, Elinor. 2002. “Against Blameless Wrongdoing.” *Ethical Theory and Moral Practice* 5: 287–303.

Mason, Elinor. 2014. “Objective and Subjective Utilitarianism’ in *The Cambridge Companion to Utilitarianism*, edited by Ben Eggleston and Dale E. Miller, CUP (2014).

(p. 178) Mason, Elinor. 2018. “Consequentialism and Moral Responsibility.” In *Consequentialism: New Directions, New Problems?* edited by Christian Seidel, 219–236. New York: Oxford University Press.

Mason, Elinor. 2019. *Ways to be Blameworthy: Rightness, Wrongness, and Responsibility*. Oxford: Oxford University Press.

McGeer, Victoria. 2014. “P. F. Strawson’s Consequentialism.” In *Oxford Studies in Agency and Responsibility: “Freedom and Resentment” at 50*, Vol. 2, edited by D. Shoemaker and N. Tognazzini, 64–92. Oxford: Oxford University Press.

McGeer, Victoria. 2015. “Building a Better Theory of Responsibility.” *Philosophical Studies* 172: 2635–2649.

Morris, Rick. 2017. “Praise, Blame, and Demandingness.” *Philosophical Studies* 17, no. 7: 1857–1869.

Oakley, Justin, and Cocking, Dean. 1994. “Consequentialism, Moral Responsibility, and the Intention/Foresight Distinction.” *Utilitas* 6, no. 2: 201.

Parfit, Derek. 1984. *Reasons and Persons.* Oxford: Clarendon.

Scanlon, Thomas. M. 1988. “The Significance of Choice”. In *The Tanner Lectures on Human Values* Vol. 8, edited by Sterling M. McMurrin, 149–216. University of Utah Press.

Scheffler, Samuel. 2004. “Doing and Allowing.” *Ethics* 114: 215–239.

Schlick, M. 1939. *The Problem of Ethics*. New York: Prentice-Hall.

Schmidt, Andreas. Unpublished manuscript. “Consequentialism and the Ethics of Blame.”

Sidgwick, Henry. 1907. *The Methods of Ethics*. 7th ed. London: Macmillan.

Smart, J. J. C. 1961. “Free-will, Praise and Blame.” *Mind* 70: 291–306.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Smith, Holly. 2010. “Subjective Rightness.” *Social Philosophy and Policy* 27, no. 2: 64–110.

Strawson, P. F. 1962. “Freedom and Resentment.” *Proceedings of the British Academy* 48: 187–211. Reprinted in Gary Watson, ed. (2003), *Free Will*, 2nd ed., Oxford: Oxford University Press, 72–93.

Vargas, Manuel. 2013. *Building Better Beings: A Theory of Moral Responsibility*. Oxford University Press.

Wallace, R. Jay. 1994. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.

Williams, Bernard. 1973. “A Critique of Utilitarianism.” In *Utilitarianism: For and Against*, edited by J. J. C. Smart and B. Williams, 77–155. Cambridge: Cambridge University Press.

Zimmerman, Michael J. 2008. *Living with Uncertainty* Cambridge: Cambridge University Press.

Notes:

([1]) Rick Morris pursues this line in a recent article (2017). See also Andreas Schmidt, “Consequentialism and the Ethics of Blame” (unpublished manuscript).

([2]) For classic critiques, see Bennett (2008), 53; Brandt (1969), 344–345, Scanlon (1988), 159–160; Wallace (1994), 4.

([3]) Strawson is sometimes interpreted as a projectivist, but that seems to me to be a different level of theorizing, akin to the distinction between normative ethics and metaethics. Strawson argues against consequentialism in normative responsibility theory, providing an account of how to assign responsibility that does not appeal directly to free will but does not appeal to consequences either. When we go to the meta level, and ask, “what makes these judgments about responsibility true?,” the best answer may be some sort of sophisticated projectivism.

([4]) Incompatibilists, of course, dispute this, and point out that if in the end the intentional injury was determined, then it is really no different from an accidental injury, in that it was ultimately caused by things originating outside the agent. My view, as I said, is that Strawson is right in seeing a difference in quality of will as crucial nonetheless, but I will leave that argument here for now.

([5]) Strawson himself makes a couple of suggestions. One is that the practice is unavoidable for us, given our psychologies; the other is that it is valuable for us. Victoria McGeer argues that we should see Strawson as a consequentialist here (McGeer 2014).

([6]) All of these theorists are interested in the issue that Brandt raises about how the moral responsibility system has consequences for our moral development. Manuel Vargas argues that the moral influence view has an over simplistic account of what conse-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

quences are relevant. We should think of the agent's development as an agent as well as about her behavior. Vargas argues that agency is an independent notion, it is more than just being influenceable—it is responsiveness to moral reasons. McGeer (2015) objects that this leaves a "justification gap": norms that are justified at the general level (e.g., norms of praising people for good will) might not apply accurately to particular cases. Thus McGeer argues for an account of agency that is not independent, according to which agency is roughly susceptibility to influence by the moral responsibility system.

([7]) Vargas is happy with this, insisting that we must accept some revisionism.

([8]) See chapters in this volume by Brad Hooker (Chapter 23) and Holly Lawford-Smith & William Tuckwell (Chapter 33) for related discussions.

([9]) I discuss blameless wrongdoing in Mason (2002).

([10]) See also Schmidt, "Consequentialism and the Ethics of Blame" (unpublished manuscript), who points out that we should understand blameless wrongdoing as a feature of sophisticated consequentialism (roughly, the view that consequentialists should think not just about acts but about what is causally upstream of acts).

([11]) The first line of the Groundwork: "Nothing in the world or out of it!—can possibly be conceived that could be called 'good' without qualification except a good will."

([12]) This is not the phenomenon of "blameless wrongdoing," which has a technical sense. I discuss "blameless wrongdoing" in section 2.

([13]) The terminology varies, but broadly there are three groups of views: objectivism, prospectivism, and subjectivism (Michael Zimmerman's terms [2008]). See my 2014 for an account of the various positions here, and my 2019 for a worked-out theory of subjective rightness. One big issue for accounts of subjective rightness is whether normative uncertainty as well as factual uncertainty should be included in the agent's point of view. I argue that subjective rightness is theory relative, so that subjective consequentialism, for example, holds fixed the consequentialist theory of value. Thus those who are (even nonculpably) ignorant of that do not count as acting subjectively rightly when they do their best by their own lights.

([14]) See Jackson's chapter (Chapter 17) in this volume for more on action guidance.

([15]) For an alternative approach to the connection between the doctrine of double effect and moral responsibility, see Oakley and Cocking (1994). They argue that agents who directly intend and agents who merely foresee have the same moral responsibility, but that permissibility is separable from moral responsibility, so the consequentialist cannot appeal to a view about moral responsibility to deny the doctrine.

([16]) See also Kamm's discussion (1999).

([17]) For a discussion of alienation and motives, see Maguire and Baker's chapter (Chapter 21) in this volume.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([18]) For criticism of Scheffler on this point, see Bradley and Stocker (2005).

Elinor Mason

Elinor Mason is Professor of Philosophy at the University of California, Santa Barbara. She works on ethics, moral responsibility, and feminist philosophy. She is the author of Ways To Be Blameworthy: Rightness, Wrongness, and Responsibility (Oxford University Press, 2019).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Population Ethics, the Mere Addition Paradox, and the Structure of Consequentialism

Melinda A. Roberts

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.20

Abstract and Keywords

In this paper, I describe three structural issues population ethics raises for any form of consequentialism that embraces what we can call the *basic maximizing idea*, the idea that it makes things better, in a morally relevant sense, to make things better for people. What we say about those structural issues will in turn determine what we say about some of the most challenging problems of population ethics. I explore a handful of our options here, discarding some and leaving others on the table. My primary focus is on how those options propose to resolve the *mere addition paradox*, a population problem that is important in its own right and whose resolution is defining for what we will want to say about many other population problems.

Keywords: population ethics, mere addition paradox, maximization, mere addition principle, Pareto plus, average view, total view, person-affecting intuition

1. Introduction

POPULATION ethics charges us with the task of comparing outcomes—that is, possible worlds or, we will say, *possible futures* or simply *futures*—the populations of which are *variable* rather than *fixed* in nature—populations, that is, that *vary* rather than remain *fixed* in respect of the overall number of people who do or will exist in the futures to be compared. The task of comparing futures containing variable populations was not completely overlooked in past decades. But just how challenging it would be to identify a form of consequentialism capable of providing a credible account of cases involving variable populations came as a surprise.[1]

Thus we find again and again that the form of consequentialism that passes one of the hard tests of population ethics seems doomed to fail another. Indeed, it's been suggested by some philosophers that a credible theory of population—a form of consequentialism

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

capable of providing a credible account of the variable population cases—simply does not exist.[2]

(p. 475) One thing seems clear. Variable population cases force us to think carefully about long-standing positions we have held and assumptions we have made regarding the very structure of consequentialism.[3]

The *mere addition paradox* is based on just one such case, and it is that case I focus on for purposes here. It's true that other variable population cases as well in one way or another raise structural issues.[4] But the mere addition paradox is *defining*: what we say about that problem informs, indeed, dictates, much of what we will say in the end about all the others.

The plan for this paper is as follows. I will start, in section 2, by laying out three structural issues that population ethics raises for any form of consequentialism that embraces what we can call the *basic maximizing idea*, the idea that it makes things morally better to make things better for people. In section 3, I describe the mere addition paradox itself. In section 4, I evaluate a handful of particularly interesting ways in which we might work to resolve our structural issues and with that work resolve that paradox. The main subject of contention in section 4 shall be the *mere addition principle*, the idea that it can't make things worse, other things equal, to add to a given future an additional person whose existence is itself worth having. Conclusions are noted in section 5.

2. The Basic Maximizing Idea and Three Structural Issues

2.1. Basic Maximizing Idea

We can't clearly have in mind any particular form of consequentialism without being clear on just what position that theory takes in respect of what we can call the *structure of morality*. Without such an understanding, the *basic maximizing idea*, that disarmingly simple idea that many consequentialists and many others as well are quick to endorse, in fact remains, on any more careful examination, hopelessly obscure.

According to the basic maximizing idea, it makes things *better* to make things *better for* people—to create, that is, *more well-being* for people rather than *less*. Consequentialists disagree whether the evaluation of possible *futures* in respect of their overall betterness is closely connected to how the *choices* that give rise to those futures are to be evaluated. If, however, a particular theory accepts that the two inquiries are closely related—accepts, that is, a strong *telic-deontic connection*—then that theory considers the basic maximizing idea to include the following practical instruction: a given choice is *wrong if* the future that obtains under that choice is *worse* than at least one alternate available or, we'll say, *accessible*, future that obtains under any alternate choice, (p. 476) and a choice

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

is *permissible* if the future that obtains under that choice is *at least as good as* any alternate accessible future that obtains under any alternate choice.[5]

The basic maximizing idea, in both its telic and its deontic forms, is highly attractive. A future that makes things better for people is better, and we ought to do the best we can. But that idea—at least, that *articulation* of that idea—is not the end of a theory but rather only the beginning.

At least three structural issues require resolution—and that's so, even if we set aside, as we do here, any number of other issues, including what counts as a *person* and what it is to make things *better for* a particular person, that is, what it is to create more *well-being* for a given person.[6] First, there is *no one way* of calculating when things are *better for people*. So we need to say *which principle of calculation* the basic maximizing idea is best understood to include. Second, there is *no one class of people* for whom things can be made better. So we need to say *which class of people* it makes things better to make things better for. Third, underlying facts regarding when one future is better for a person than another—when a person has more well-being rather than less—can be sorted into different classes. So we need to say *which facts relating to a future's being better for a person—which betterness-for facts*—make things better.

2.2. Which Way to Calculate When Things Are Better for People?

It may not be immediately clear in each case just what the structural issue is, and so I will say more about all three in this section and the two sections that follow.

(p. 477) The debate surrounding the first structural issue, the issue of how we are to calculate betterness for people, will be familiar.[7] One way to make things better for people is to make things better in *total*. Another way to make things better for people is to make things better on *average*. Traditionally, the *total* principle and the *average* principle both commence by taking the summation of the raw, unadjusted well-being levels of all the people who do or will exist in that future.[8] Under the total principle, that summation itself just is the overall value of the future. Under the average principle, the overall value of the future consists in that same summation but divided by the number of people who do or will exist in that future.

Both the total principle and the average principle are *additive*, or *aggregative*, in nature. And they generate the same overall betterness results—the same *ranking* of possible futures in respect of their overall betterness—in any *fixed* population case. But they generate different overall betterness results—different rankings—in many *variable* population cases. According to the total principle, additional people who have existences worth having make a future better *provided that* their existence doesn't make things *too* much worse for anyone else. Averagism is not so quick to say that additional worth-having existences make a future better. If adding more people means that *average* well-being will decline—if, for example, the larger population means that resources and thus well-being are

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

spread too thin—then, according to averagism, the future that includes those additional people will be worse rather than better.

A third way of calculating how things are made better for people—a way that may be attractive to theorists who object to the accounts the total principle and the average principle give of cases that involve *tradeoffs* between people and, specifically, object to results those theories generate in cases in which some people must suffer horribly so that other people can have their wonderful lives—is *nonaggregative* in nature. According to that third principle—what we can call the *quantificational* principle—to make things better for people is to bring it about that *for each person* in that class, for *each person* as an individual, things are made better *for that person*.

In contrast to the total principle and the average principle, the quantificational principle doesn't, on its own, determine how tradeoffs are to be made. A reasonably complete theory that determines value by reference to the quantificational principle of course will need to say how that is to be done. However, an advantage of the principle is that it *doesn't* commit us in advance to the position that the tradeoff between, for example, many people suffering mosquito bites and one person being subjected to intense torture is to be settled in favor of the one person's being tortured.[9]

(p. 478) 2.3. Which Class of People Does It Make Things Better to Make Things Better for?

As just noted, the total principle and the average principle traditionally calculate the value of a future by reference to the well-being levels of all the people who *do or will exist* in that future. But that traditional approach is not necessarily implicated by those principles. A second structural issue arises out of that point: we will need to say *which class of people* it makes things *better* to make things *better for*. Is it *everyone* who does or will exist in the relevant future? Do *all* those people *matter morally*? Do they *all* have *moral status*? Or not?

That second structural issue arises with particular clarity in the variable population cases. Graph 25.1 describes such a case. Here, we are asked to compare two futures, the second of which includes people who never exist at all in the first.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

	c1	c2
	f1 (actual)	f2
+15	**p1 . . . p10**	
+10		**p1 . . . p10,** **q1 . . . q10**
+5		
+0	*q1* . . . q10**	

Graph 25.1 Actual People Versus Merely Possible People

In this graph and the graphs that follow, c1 and c2 represent alternate choices, and f1 and f2 represent alternate accessible futures that obtain, respectively, under those choices. The alternate accessible futures and the alternative choices involved in the case are to be understood as exclusive and as exhausted by the graph. A person's name in bold means that person does or will exist in the indicated future, and a person's name in italics with an asterisk (*) means that person never exists in the indicated future. It's a stipulation of the case that the choice of c1 is in fact made and that the future f1 in fact—that is, *actually*—unfolds. Thus the *actual* people, p1–p10, exist in both f1 and f2, while the people who are *merely possible* relative to f1, q1–q10, exist only in f2. Levels of well-being for each person at each accessible future are indicated in the far-left column. It's an assumption for purposes here that a person's well-being level at any world where that person never exists is just zero.[10]

(p. 479) We can now articulate the second structural issue. Which *class of people* are things to be made better for? On an *inclusive* view (let's call it *moral possibilism*) the class of people things are to be made better for consists of *all* the people who exist in *any* of the alternate accessible futures—the class, that is, consisting of all of p1–p10 and all of q1–q10; of all *actual* people as well as all *merely possible* people.

By tradition, both the theory that adopts the total principle and the theory that adopts the average principle also adopt moral possibilism. But we can easily see that the two theories provide inconsistent accounts of the case. The totalist account implies that f2 is better than f1 while the averagist account implies that f1 is better than f2.

But neither the totalist nor the averagist *must* accept moral possibilism. Instead, taking the position that *moral* status is determined by *existential* status, they might adopt a more *exclusive* view. They might, that is, adopt *moral actualism*, according to which it's only *actual* people who have moral status. The people we make things better by making things better for are, in other words, all and only *actual* people—people, that is, who do or will exist in the uniquely *actual* future.[11]

Moral actualism represents one way—but, as we shall see, not the only way—of understanding Narveson's often cited idea that we should be interested not in "making happy

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

people" but rather in "making people happy."[12] Narveson's idea is highly intuitive. If moral actualism is considered to capture that idea, then that fact may itself explain why moral actualism has seemed so attractive to some philosophers.[13]

Under moral actualism, the totalist account and the averagist account produce identical results. Both instruct that f1 is better than f2 in virtue of the fact that, under both the total principle and the average principle, f1 is better for the only people who matter morally, that is, *actual* people.

The issue of *which people* we are to make things better for—the issue of who *matters morally*; of who has *moral status*—also arises under the quantificational principle. On that principle, what makes things better for people is that things be made for people on a quantificational basis. That principle in combination with moral possibilism suggests that it's just as important to make things better for merely possible people as it is to make things better for actual people. Since the case at hand involves a tradeoff and, as noted earlier, the quantificational principle can't on its own tell us how any such tradeoff is to be made, we of course can't here complete the quantificational account of the case.

(p. 480) In contrast, under moral actualism, we need only worry about making things better for actual people. If that's our view, then the quantificational principle will, like the total principle and the average principle, easily generate the result that f1 is better than f2.

2.4. Which Betterness-For Facts Have Moral Significance?

The third structural issue—the issue of when the fact that one future is better for a person than another future has moral significance; that is, which *betterness-for* facts have moral significance—is just as critical as the first two.

In many cases, it will be clear whether the basic maximizing idea deems a certain betterness-for fact to have moral significance. For example, it's clear that the basic maximizing idea, for purposes of determining which future is better, will take into account the fact that the future in which a happy person's worth-living life *isn't* prematurely ended by murder is better for that person than the future in which that same happy person is allowed to live life out to its natural conclusion. In that case, the person the second future makes things better for exists in *both* futures.

But in other cases the person the betterness-for fact is a fact *about* exists in only one of the two futures to be compared. If a person never exists in one future but has an existence worth having in another future, then it follows that the other future is better for that person than the one.[14] But the question whether that particular betterness-for fact has *moral significance* will remain open.

Thus the third structural issue: do *both* sorts of betterness-for facts—the perfectly *ordinary* as well as the *existential*—have moral significance? Or is it just the former?

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

On some ways of resolving our first two structural issues, what we should say about the third structural issue may seem automatically settled. Thus it might be hard to see how the theorist who accepts *both* the total principle *and* moral possibilism can deny that all betterness-for facts have moral significance.

But that that's so is not a foregone conclusion. Broome, for example, an avowed totalist, outlines a nonstandard theory that may have the resources to do just that.[15] Specifically, what Broome's theory adds up for the purpose of determining the overall value of a given future doesn't consist of raw, unadjusted *well-being* levels but rather what Broome calls levels of the *personal good*. As Broome understands it, the concept of the personal good is malleable enough to reflect not just raw well-being—the value, that is, a given future has for a given person—but also such values as equality, fairness, and priority. I have elsewhere proposed that we might extend Broome's concept still further and understand it to reflect certain *existential* values as well, values that, in effect, temper the *maximizing* values we, as consequentialists, find so compelling.[16]

(p. 481) Still, as a matter of tradition, the total principle and the average principle go hand in hand with a *broad* understanding of just which betterness-for facts have moral significance—a *wide betterness-for principle*. Thus the totalist and the averagist will, as a matter of tradition, take the position that both ordinary and existential betterness-for facts have moral significance.

The third structural issue should be viewed as particularly critical for the quantificational theorist who finds Narveson's thought that we should be interested not in "making happy people" but rather in "making people happy" highly intuitive *but also* finds moral possibilism compelling. It would be a mistake, that is, for such a theorist to follow the lead of most consequentialists and take the wide betterness-for principle for granted. For, as we will see in what follows, moral possibilism *is* compelling.[17] The solution is for the theorist who wants to retain the Narvesonian thought to accept that *all* people matter morally but insist that *not* all betterness-for facts have moral significance. A *narrow betterness-for principle*—one that provides that the fact that one future is better for a person than an alternate future has moral significance *only if* that person does or will exist in the alternate future—offers a way to achieve that latter aim without abandoning the former.[18]

To see how the narrow betterness-for principle works, consider the *procreative asymmetry*. According to the asymmetry, it makes things *worse*, other things equal, to bring a *miserable* child into existence but *doesn't*, other things equal, make things *better* to bring a *happy* child into existence.[19] On the assumption that existence is worse than nonexistence for the miserable child and better for the happy child, the narrow principle instructs that making things better for the miserable child by leaving that child out of existence—regardless of whether that child *actually* exists—is a betterness-for fact that has moral significance.[20] Leaving that child out of existence counts, we can say, *in favor of* the future in which that child never exists at all, while bringing that child into existence counts *against* the future in which that child exists. In contrast, according to the narrow principle, making things better for the happy child by bringing that child into existence is

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

a betterness-for fact that has *no* moral significance since the future in which that child is worse off is a future in which that child never exists at all.

The narrow betterness-for principle thus offers a way of understanding Narveson's highly intuitive idea—that we make things better not by "making happy people" but by "making people happy"—that is perfectly consistent with moral possibilism—that doesn't, that is, presuppose moral actualism. The principle itself is itself plausibly rooted (p. 482) in what Parfit called the *person-affecting intuition* and I will here call the *person-based intuition* (*PBI*): "what is *bad* must be bad *for someone*."[21] Unpacking Parfit's pithy version just a bit, we can articulate PBI as follows: the *worse* future, and the *wrong* choice made at that future, must make things *worse for* at least one person who *does or will exist* in that future. Now, this formulation of PBI is itself elliptical (the one future is worse for the one person *than what*?), and we will need to return to how that ellipsis is to be completed in what follows.[22] But even in its elliptical form PBI suggests no basis for thinking that PBI *requires* moral actualism (or vice versa). It can equally well instead be understood in terms of the narrow betterness-for principle.

3. The Mere Addition Paradox

Section 2 outlined three structural issues we must resolve to fix the content of the idea that what makes things better is to make things better for people—of, that is, the basic maximizing idea. Depending on how we resolve those three issues, we generate different versions of the basic maximizing idea.

The mere addition paradox raises those structural issues in all their complexity. We thus turn to that problem now.

3.1. The Mere Addition Case and the All-Critical Mere Addition Principle (MAP)

Graph 25.2 (the *Mere Addition Case*) describes the case that generates paradox. Graph 25.2 is to be understood just as Graph 25.1, except that I use Parfit's now familiar terminology to designate the possible futures that are to be compared and I add corresponding designations for the choices that give rise to those futures.[23]

choices	c(A)	c(A+)	c(B)
futures	A	A+	B
+15	p1 . . . pn	p1 . . . pn	
+12			p1 . . . pn, q1 . . . qn
+1		q1 . . . qn	
+0	*q1* . . . qn**		

Graph 25.2 Mere Addition Case

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

We can now state the paradox.[24] We start, according to Parfit, by agreeing that B is worse than A. For the claim that B is worse than A holds, he notes, under various alternative views regarding value that he finds plausible.[25] But it's perhaps not just those (p. 483) various alternative views that made Parfit—and makes us—want to say that B is worse than A. Parfit may also have been thinking of his own infamous *repugnant conclusion* as constituting, in effect, an argument that B is worse than A.[26] We'll return to the repugnant conclusion later. For purposes here, it's enough to say that the futures Parfit has us consider in that case include the future A, in which a very large number of people have lives that are well worth living, and Z, in which a much larger number of people have lives that are *just barely* worth living. As Parfit notes, the conclusion that Z is better than—or even that it's at least as good as—A seems repugnant; that is, it is clearly false.[27]

Thus line (1) of the proof, itself a premise: that B is worse than A.

We now compare A against A+. It is at this point that what is perhaps the most quietly controversial of all the principles that have helped shape population ethics over the last few decades is put to work—that is, the seemingly unassuming *mere addition principle* (*MAP*).

> *Mere addition principle (MAP)*. Where a future x includes exactly the same people as an alternate future y, with each person in x having exactly the same well-being level as that person has in y, *except* that an additional person exists in y and has an existence worth having in y, it's not the case that y is worse than x.

Mere addition, as Parfit puts it, happens when the additional people have "lives worth living" and their existences "affect no one else" and involve no "social injustice."[28] When those conditions are met, MAP claims that the existence of the additional people doesn't make things worse. How, as Parfit himself put it, can *mere addition* make things *worse*?

Applied to the mere addition case, MAP implies that it's not the case that A+ is worse than A. Thus line (2) of the proof: that A+ isn't worse than A.

We now compare A+ against B. Here we face a fixed population comparison; A+ and B, that is, contain exactly the same people. Again citing various alternative views regarding value that he finds plausible, Parfit concludes that B is surely better than A+.[29]

(p. 484) That result may stand as among the least controversial claims that emerges in connection with the mere addition paradox. Thus line (3) of the proof: that B is better than A+.

We now consider what we can infer from lines (2) and (3) together. Since A+ isn't worse than A and B is better than A+, it seems to follow that B isn't worse than A. ("B cannot be worse than A if it is better than something—A+—which is not worse than A."[30]) And thus line (4) of the proof: that B isn't worse than A. Lines (1) and (4) together produce a contradiction—that is, line (5) of the proof.

The proof can be summed up as follows:

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Mere Addition Paradox

1. B is worse than A (premise; implication of independent argument)

2. It's not the case that A+ is worse than A (MAP)

3. B is better than A+ (premise; implication of independent argument)

4. It's not the case that B is worse than A (from (2) and (3))

Therefore:

5. Contradiction ((1) and (4))

There are gaps in this line of reasoning. The argument can't cleanly get—via the principles of first-order logic alone—from (2) and (3) to (4) without implicitly endorsing certain conceptual principles. Those include the following: for any futures x and y, x is worse than y, y is worse than x, or x is exactly as good as y; if x is better than y, then it's not the case that x is worse than y; and, for any futures x, y, and z, if x is exactly as good as y and z is better than x, then z is better than y. The argument further assumes—and this is an assumption we shall revisit below—*transitivity*, that is, the principle that, for any futures x, y, and z, if x is better than y and y is better than z, then x is better than z.

Those principles in hand, we can provide a subproof in favor of (4), as follows:

(i) Either A+ is exactly as good as A or A+ is better than A (conceptual principle, line (2))

(ii) If A+ is exactly as good as A, then B is better than A (conceptual principle, line (3))

(iii) If A+ is better than A, then B is better than A (transitivity, line (3))

(iv) B is better than A (constructive dilemma, lines (i)–(iii))

(v) If B is better than A, then it's not the case that B is worse than A (conceptual principle)

(4) It's not the case that B is worse than A ((iv) and (v)).

And (4), as noted earlier, contradicts (1), producing (5), a contradiction.

We thus face a bona fide *paradox*—a valid argument that relies on premises and assumptions that themselves seem compelling.

(p. 485) How might we resolve that paradox? As we shall see in what follows, there is no easy answer to that question, with each proposed resolution seeming to generate its own set of puzzles and problems.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

4. Options for Resolving the Mere Addition Paradox

4.1. Rejecting Premise That A+ Isn't Worse Than A

Let's start big: let's first take on the all-critical *mere addition principle* itself. If we reject MAP, we can reject line (2) of the mere addition paradox and easily avoid contradiction.

4.1.1. MAP and Pareto Plus

MAP itself has some surprisingly strong implications—*controversial* implications that give us a reason to consider rejecting that principle from the start—or at least to understand that it's not a principle we can legitimately *assume*.

Thus MAP itself may seem not to say much. Consider what MAP *doesn't* claim: it *doesn't* claim that A+ is *better* than A. In contrast, the principle we can call *Pareto plus* makes exactly that claim.[31]

> *Pareto plus*. Where a future x includes exactly the same people as an alternate accessible future y and each person in x has exactly the same well-being level as that (same) person has in y *except* that an additional person exists in y and has an existence worth having in y, then y is *better* than x; that is, x is worse than y.

Parfit—fascinatingly—says much less about Pareto plus than we might expect. That may have been a strategic choice. He may have, that is, realized that Pareto plus, in contrast to the seemingly unassuming MAP, would be considered highly controversial from the start.

Why? What reasons might we have to think that Pareto plus should be considered controversial? Here are two. First, to accept Pareto plus—in the absence of argument—would seem to beg some of the structural questions noted in section 2. Consider, for example, the narrow betterness-for principle, itself closely related to PBI. That principle implies that A+ *isn't* better than A—that the additional lives worth living in A+ *don't* count in favor of A+ or against A. To assume Pareto plus when the truth of the narrow betterness-for principle itself is what we are trying to resolve would be to beg the question against that principle.

(p. 486) Should the fact that Pareto plus rules out the narrow betterness-for principle and thus PBI concern us? Perhaps not. Perhaps we, instead, should view Pareto plus as offering a helpful resolution of that particular structural issue.[32]

However, we have another reason as well to consider Pareto plus controversial. Assume that the mere addition case itself is one in which the telic-deontic connection holds. When we then say that A+ is better than A, we are at the same time saying that the choice that gives rise to A—that is, c(A)—is *wrong*. That pro-procreation result might seem perfectly plausible in an other-things-equal case, including the case at hand. But if we accept Pare-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

to plus, then it seems that, at that point where we decide how to situate Pareto plus in a broader theory, we'll also find ourselves committed to accept a certain *extension* of Pareto plus. If the addition of the worth-having existence makes things better when other things are equal—when it comes at *no* cost to anyone else—then it seems that the addition of the worth-having existence also makes things better when it comes at *some* cost to others. But do we really think that a couple who is thinking of having a third child and who understands that their doing so will mean that certain needs of their already-existing children will go unmet is wrong *not* to produce that third child?

Most of us probably think they are not. But whatever our instincts about that case happen to be, the following seems clear: the implications of an extended Pareto plus, in combination with the principle of connection, will be highly controversial and require further discussion. I won't try to complete that discussion here. For purposes here, we should just note that we have not one but two reasons for considering Pareto plus controversial. It's not a principle we are compelled, starting out, to accept or entitled to assume.

What I now want to show is that, while Pareto plus seems far more controversial than the unassuming MAP, the fact is that, if we accept MAP, we are forced to accept Pareto plus as well. That means that, if Pareto plus is controversial—is, that is, not a principle we are entitled from the start to assume—then so is MAP. And if MAP is controversial—if it's just as controversial and potentially question-begging as Pareto plus—then rejecting MAP as a strategy for resolving the paradox should be seriously considered.

My argument is based on a line of reasoning articulated by John Broome in another context. Consider Broome's *Three Outcome Case* (Graph 25.3).

choices	c1	c2	c3
futures	f1	f2	f3
+15	p1 . . . pn	p1 . . . pn	Harry, p1 . . . pn
+10		Harry	
+0	*Harry**		

Graph 25.3 Three Outcome Case

In this case, Harry never exists at all in f1. He has a life well worth living in f2 and a substantially better life in f3. Finally, Harry's addition in both f2 and f3 meets the conditions for *mere* addition. Thus, among other things, the members of the overlapping population p1 … pn are left unaffected as a result of Harry's coming into existence in f2 and f3.

We commence the argument by assuming that MAP itself is true. We assume, as a further conceptual principle, that if a future x isn't worse than a future y, x is at least as good as y. We then infer from MAP that each of f2 and f3 is at least as good as f1. We also (p. 487) assume that f1 is at least as good as f3. (It's not plausible to say that the worth-having existence of the additional person *always* makes things worse; moreover, in the case at

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

hand, Harry's well-being is *maximized* in f3. If ever such an addition *doesn't* make things worse, it will be in just such a case.) We then infer that f1 is exactly as good as f3. An uncontroversial, same-people Pareto principle tells us that, since f2 and f3 contain exactly the same people, and f3 is better than f2 for Harry and worse for no one, f3 is better than f2. Together, those claims establish that f1 is better than f2—that is, that f2 is worse than f1. Let's now assume, for purposes of *reductio*, that Pareto plus is false—that it's *not* the case that f2 is better than f1; that is, that f2 is *at least as good as* f1. We immediately have a contradiction—f2 can't be both at least as good as f1 *and* worse than f1. Completing the *reductio*, we say that our assumption must have been false—that Pareto plus, after all, is true. And that, in turn, completes the conditional proof: if MAP is true, then so is Pareto plus.[33]

I have not tried to argue here that MAP, on its own, without reference to any other conceptual principles or logical truths, generates Pareto plus. What I have argued, rather, is that MAP, in combination with some minimal, highly plausible assumptions, generates Pareto plus. That I think is a surprising result—at least, it surprised me.

Now, consistent with our conditional conclusion, we of course remain perfectly free to reject *both* MAP *and* Pareto plus. If we then decide at the end of the day that Pareto plus must go—that it's too controversial to help us resolve our structural issues—then we should decide that MAP must go as well. And that decision, in turn, opens the door to a way of resolving the mere addition paradox: we reject MAP and, with MAP, line (2) of the paradox, that is, the claim that A+ isn't worse than A.

4.1.2. Averagism as a Basis for Rejecting MAP

The position that we can resolve the mere addition paradox by rejecting MAP challenges us to identify a plausible theory on which MAP is false—a theory that resolves the three structural issues described in section 2 in a way that rejects MAP and thereby resolves the mere addition paradox.

(p. 488) The total principle, in combination with moral possibilism and the wide betterness-for principle—that is, the traditional total view; what we will call *totalism* going forward—won't, of course, do for that purpose since MAP is simply a theorem of totalism.[34]

The average principle, in combination with moral possibilism and the wide principle—that is, the traditional average view; what we will call *averagism*—directly challenges MAP. Averagism immediately generates the result that A+ is worse than A, thereby rejecting MAP. Moreover, averagism completes the account of the mere addition case in a way that seems clearly plausible. Thus averagism instructs that A+ is worse than A, that B is worse than A, and, finally, that B is better than A+.

But the strategy of using averagism to justify the rejection of MAP is highly problematic. For averagism is itself highly problematic. One of Parfit's arguments against averagism seems particularly hard to resist. Consider Graph 25.4 (*Hell Three*).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

choices	c1	c2
futures	f1	f2
+0	$q1^* \ldots qm^*$	
−100		q1 . . . qm
−1000	p1 . . . pn	p1 . . . pn

Graph 25.4 Hell ree

Graph 25.4 below will add some detail to Parfit's own case. But Parfit's underlying point retains its full force. Averagism implies that f2 is better than f1. And, as Parfit writes, the implication from averagism that "we ought to have these children"—that is, that we ought to bring q1 … qm into a miserable existence since doing so raises the average well-being level—seems clearly false. "We have seen enough to know that we should reject this principle."[35]

4.1.3. Person-Based Approach as a Basis for Rejecting MAP

Let's call the view that accepts the narrow betterness-for principle—and, with that principle, the person-based intuition itself, or PBI—as well as moral possibilism the *person-based view*. I'll take it for granted that the person-based view is quantificational in nature. It will decide whether a future is better for people by looking at each person as (p. 489) an individual and not by reference to the total or average principle, both of which are additive.[36]

We can see how the person-based view provides a basis for rejecting MAP by taking another look at the Three Outcome Case (Graph 25.3). Here, we assume, as before, the same-people Pareto principle as well as certain conceptual principles, including the principle that, if it's not the case x is worse than y, then x is at least as good as y, and the principle that, if x is at least as good as y and y is at least as good as x, then x is exactly as good as y.

The person-based account of that case is straightforward. PBI implies that f1 is at least as good as f2. PBI also implies that f1 is at least as good as f3 and vice versa. Those results, in combination with our conceptual principles, tells us that f1 is exactly as good as f3. The same-people Pareto principle, in addition, implies that f2 is worse than f3. But if f1 *is exactly as good as* f3, and f2 is *worse than* f3, we can then infer that f2 is also *worse than* f1.

That, in turn, tells us that MAP is false. We thus seem to have, in the person-based view, a basis for rejecting MAP.

However, if, in another person-based inference or two, we find ourselves enmeshed in inconsistency, then the person-based view will clearly fail as a basis for rejecting MAP. Consider how PBI compares f1 and f2. Harry *exists* in f2, but f2 can't plausibly be considered *worse* for him than f1. Isn't PBI's necessary condition on worseness thus left unsatisfied,

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

forcing us to conclude that, after all, f2 is *at least as good as* f1? Don't we then face inconsistency, having already concluded that f2 is *worse than* f1?

Not necessarily. It depends on how we complete the ellipsis in PBI—in effect, on what betterness-for facts we take to have moral significance.

Thus, according to PBI, for the one future to be worse than a second, the one future must be *worse for* a person who does or will exist in the one future. But worse for that person *than what*? We have two options for answering that question. We can take a *restrictive* view and say that, according to PBI, x is worse than y *only if* there does or will exist a person p in x such that x is worse for p *than y*.[37] Or we can take an *expansive* view and say that, according to PBI, x is worse than y *only if* there does or will exist a person p in x such that x is worse for p *than z*, where z is an alternate accessible future relative to x, z may but need not be identical to y, and x is worse for p *than z*.[38] We can accept a (p. 490) *restricted* class of betterness-for facts as morally significant, or we can accept an *expansive* class of betterness-for facts for that purpose.

Clearly, for the person-based view to remain plausible, we must adopt the *expansive* view. We thus complete the ellipsis as follows.

> *Expansive person-based intuition* (*expansive PBI*): x is worse than y *only if* there is a person p and an alternate accessible future z such that p does or will exist in x and x is worse for p than z (where z may, but need not, be identical to y); and
>
> a choice c made at x is wrong, *only if* there is a person p and an alternate choice c′ at an alternate accessible future y such that p does or will exist in x and x is worse for p than y.

According to expansive PBI, f2 is worse than f1 *only if* there is someone in f2 such that f2 is worse for that person than is *some* alternate accessible future. Since f2 is worse for Harry than *f3*, the necessary condition is satisfied. We thus avoid the implication that f2 is at least as good as f1 and remain free to say—and *do* say, as noted earlier—that f2 is worse than f1.

We now have a complete and consistent person-based account of the Three Outcome Case—and a basis for rejecting MAP.[39] It rejects MAP—and accepts, in its place, what we can call *Pareto minus*: on occasion, the addition of the worth-having existence, other things equal, makes things *worse*.

As a basis for rejecting MAP, the person-based view will nonetheless remain controversial. A major stumbling block has been the *nonidentity problem*.[40] The objection there is that the person-based view—and PBI specifically—implies that it doesn't make things worse, and is perfectly permissible, to make choices that will impose great burdens on future people in cases in which clearly better alternatives are available. Consider, for example, the choice of the depletion of natural resources over their conservation, or the choice of the risky policy with respect to the disposal of nuclear waste over a safe policy. Accord-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ing to the nonidentity problem, PBI implies that those choices don't make things worse—that they are perfectly permissible. But clearly they're not.

But why think that PBI in fact generates those false results? The idea there is that the choice of depletion, or the risky policy, can't make things worse for future people since any such future person can be understood on closer inspection to *owe his or her very existence* to that choice having been made. How likely is it, Parfit asks, that any of us would have existed had the world wars never have been fought, or had "motor cars" never have existed?[41] *Any* little change in the history of the world that ended in your (p. 491) coming into existence or my coming into existence could easily have thrown us off track for ever coming into existence at all, our own coming into existence in each case being highly precarious, that is, highly sensitive to the "timing and manner" not just of our own conception but that of our parents, grandparents, and indeed all our forebears, whatever their species, going back millions of years.[42]

However, whether PBI can provide a credible account of the nonidentity problem is a matter of ongoing debate. I have elsewhere argued that an *expansive* PBI in fact nicely avoids many of the results attributed to it, including that depletion doesn't make things worse and isn't wrong.[43] The idea there is that our coming into existence is highly precarious whether agents choose depletion or conservation—that under either choice the probability that any one future person will ever exist at all is equally and unbelievably small. I won't, however, try to provide any full defense of PBI here. For purposes here, the important point is just that, if the question can indeed be cleared up in favor of a person-based view, then that view will, after all, provide an intuitive basis for rejecting MAP and thus for resolving the mere addition paradox.

4.2. Rejecting Premise That B Is Worse Than A

4.2.1. Totalism as a Basis for Rejecting Claim That B Is Worse Than A

A second option for resolving the paradox is to retain MAP but reject the claim that B is worse than A. This would be to reject line (1) of the mere addition paradox. Thus we take the position that B—which itself includes *double* the number of people existing in A, *all* living lives that are *unambiguously* worth living—is *better* than A. We can further—and plausibly—claim that B is better than A+. And we can finally—and here our appeal might be to Pareto plus—claim that A+ is better than A. Those claims together constitute a complete and consistent account of the case.

The theory most closely associated with such an account of the case is *totalism*—that is, the total principle, in combination with moral possibilism and the wide betterness-for principle. A difficulty for the totalist account, however, is underlined in the proposed resolution itself. Is it really plausible to say that B is better than A? As Parfit himself has demonstrated, when we apply the line of reasoning that instructs that B is better than A to still other cases, we seem to obtain clearly false results.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Consider, for example, Parfit's repugnant conclusion case.[44] There, we are to compare a future A that includes a very large population of people all of whom have lives well worth living against a future Z that includes a much larger population of people all of whom have lives only barely worth living. To make the case even more of a challenge for the totalist, let's assume that the populations overlap: that everyone who exists in the first future also exists in the second. The case, then, can easily be detailed in such a way that (p. 492) the summation of the individual well-being levels in Z is greater than the summation of the individual well-being levels in A. Totalism, on those facts, seems immediately to instruct that Z is better than A. But that result seems clearly false—or indeed, as Parfit himself puts it, *repugnant*.

That we reject the claim that Z is better than A does not, however, mark the end of the discussion of whether totalism can provide a plausible resolution of the mere addition paradox. For one thing, totalism itself can be formulated in different ways—with some of those formulations specifically designed to avoid at least some versions of the repugnant conclusion.[45] Moreover, some philosophers have made the argument that the repugnant conclusion is not really so repugnant after all.[46]

Totalism thus remains a potentially viable, if highly controversial, basis for rejecting the claim that B is worse than A and thus for resolving the mere addition paradox.

4.2.2. Pareto Plus and the Moral Status of the Merely Possible People

As we have just seen, totalism faces a serious challenge in the form of the repugnant conclusion. On the other hand, totalism has its advantages. The fact that totalism implies Pareto plus counts, in the minds of at least some philosophers, as just such an advantage.

Specifically, Ng considers the fact that Pareto plus forces us to "consider the possible welfare of the prospective people" makes it an important safeguard against some very serious pitfalls.[47]

(p. 493) On Ng's view, the immediate basis for the claim that *mere addition* makes things better is not that aggregate well-being has been increased but rather that "no one has more moral right than another."[48] *If* the maximization of well-being in the aggregate makes things better, the *reason* that it makes things better is that prospective people—people who may, but need not ever, exist—have claims that we must "'honour'" alongside the claims of existing and future people.[49] "[T]hey are all unborn now, they are all prospective people; no one has more moral right than another."[50] "Clearly, the superiority of A+ over A … is … very compelling."[51]

Ng's view that "prospective people"—a class that would include the merely possible—have moral rights may seem controversial. As noted earlier, however, and as we shall shortly see, moral possibilism is a view we are compelled to accept and a view we shall *want* to accept once we understand the stakes.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

However, the next step of Ng's analysis is more controversial. Ng suggests that the moral rights held by people who may, but need not ever, exist include the right, other things equal, to be brought into existence. But here Ng may be conflating two distinct structural issues. One is the issue of *who* matters morally, and the other is the issue of *what it is* for a person to matter morally—what betterness-for facts have moral significance. We can thus agree with Ng that prospective people—including the merely possible—matter morally but insist that not all betterness-for facts are themselves morally significant. Thus we can say that all possible people—you, me, and the merely possible—matter morally but that none of us have the right to be brought into existence. We can, in other words, think of ourselves *and* the merely possible as having *other* "rights"—Harry, for example, in the Three Outcome Case may have the "right" against having the sort of existence he has in f2, given that he could have had, at no cost to anyone else, the sort of existence he has in f3—but not the "right" to be brought into existence.

4.3. Rejecting Premise That B Is Better Than A+

The person-based and totalist views outlined in sections 4.1 and 4.2, respectively, both support or are at least consistent with the highly plausible claim that B is better than A+—that is, line (3) of the mere addition paradox. We now ask whether there are grounds for thinking that claim is false.

Perhaps the most credible argument in favor of the position that B isn't better than A+ derives from the claim that for purposes of resolving the mere addition paradox the only people we need to worry about—the only people who matter morally—are the people (p. 494) who exist in A. Since those people are better off in A+ than they are in B, we should, accordingly, take the position that A+ is better than B. We thus avoid the paradox.

That claim is itself most plausibly rooted in moral actualism, the idea that the only people who matter morally are the people who do or will *actually* exist. Taking that position, and assuming that A itself is the actual future, we can then say that the people who exist in both A+ and B but don't exist in A have no moral status. Whether we adopt the average or the total principle—or even a quantificational principle—will not change our result given the facts of the particular case. Moral actualism, on its own, tells us that the merely possible do not matter morally. It thus easily supports the position that the fact that the merely possible are worse off in A+ than in B is itself without moral significance, leaving us free to say, in turn, that A+ is better than B.

Moral actualism is, however, highly problematic. For one thing, it's an approach that avoids the paradox only in the case where the assumption that A is the actual future holds. For another, moral actualism violates Rabinowicz's principle of normative invariance, according to which the better future—and, we can add, the permissible choice—cannot vary depending on which future itself happens actually to unfold.[52] What actually happens cannot determine what ought to have happened.[53]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

In contrast, both the person-based view and totalism are understood for purposes here to include moral possibilism, the principle that all people matter morally—you, me, and the merely possible. We each have the same moral status as anyone else.

Where the person-based view and totalism differ is on the question of what it is for a person to have moral status. Consistent with the total view—and Pareto plus—is the idea that moral status means that it would make things worse, and be wrong, other things equal, to leave any of us—you, me, or the merely possible—out of an existence worth having. Consistent with the person-based view is the idea that moral status doesn't go that far: to have moral status means that it would make things worse for me or you or the merely possible to exist, now or later, in a given future and be less well off in that future than we are in an alternate accessible future. But it doesn't mean that it would make things worse, other things equal, never to have brought us into existence to begin with.

4.4. Rejecting Transitivity

The mere addition paradox, as noted earlier, relies on the principle of transitivity. According to that principle, if x, y, and z are alternate futures, then, if x is at least as good as y and y is at least as good as z, then x is at least as good as z.

If we can plausibly deny transitivity, then we can resolve the mere addition paradox—and at the same time come to a new understanding of the nature of morality.

(p. 495) Perhaps the strongest argument against transitivity has been provided by Larry Temkin.[54] On his view, we can accept that A+ is at least as good as A and that B is better than A+ but nonetheless reject the claim that B is better than A. We thereby avoid the paradox.

But on what basis can we reject transitivity, a seemingly necessary conceptual truth? According to Temkin, in evaluating the case, we should not only compare the overall values (e.g., aggregate well-being) of each of the three futures calculated in isolation from one another but also compare the futures in terms of still other values that may—or may not—become relevant depending on which pair of futures is itself under comparison. A simple example: when we compare A against A+ under the latter approach—what Temkin calls the *essentially comparative approach*—the inequality we see in A+ is irrelevant to the comparison and the correct evaluation is that A+ is at least as good as A. But when we compare A+ against B, the inequality we see in A+ becomes highly relevant and helps to support the result that B is better than A+.[55]

If the essentially comparative view is a component of morality, then it's not surprising, Temkin argues, that transitivity fails.[56]

Temkin's complex picture of morality is both intuitive and at the same time complex. But it remains controversial. It may thus seem that we can't be confident, as a conceptual matter, that we've properly understood what all-things-considered *betterness* consists in and, at the same time, *reject* transitivity. Moreover, we may be hesitant to accept

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Temkin's pluralism at least at this point in its development. It tells us that both the essentially comparative approach and the internal aspects approach are important but not how they are to be reconciled against each other.

But we should remind ourselves that each of the other proposed resolutions of the paradox we've considered here faces certain challenges as well. Thus Temkin's perfectly cogent proposal should be considered a viable, if controversial, way of resolving the mere addition paradox.

5. Conclusion

This paper lays out three structural issues that we face when we try to give content to the basic maximizing idea, the idea that making things morally better is a matter of making things better for people. Decisions on those issues, in turn, help to define some of the (p. 496) options that we have as we work to resolve the mere addition paradox. This paper explores and begins to evaluate a handful of those options. It doesn't explore all of them, nor does it explore other interesting and important proposed resolutions of the paradox that have little or nothing to do with the three structural issues identified here.[57] Of the options the paper does explore, some are set aside and some are left on the graph. What is clear is that more work in the riveting area of population ethics—and specifically in understanding the basic structure of consequentialism itself—is urgently required. It's not just population ethics that is at stake. It is the whole of consequentialism.

References

Arrhenius, Gustaf. 2000. "An Impossibility Theorem for Welfarist Axiology." *Economics and Philosophy* 16:247–266.

Bader, Ralf. forthcoming. "Person-Affecting Utilitarianism." In *Oxford Handbook of Population Ethics*, edited by G. Arrhenius, K. Bykvist and T. Campbell. Oxford University Press.

Broome, John. 2004. *Weighing Lives*. Oxford University Press.

Broome, John. 2015. "General and Personal Good: Harsanyi's Contribution to the Theory of Value." In *The Oxford Handbook of Value Theory*, edited by Iwao Hirose and Jonas Olson, 249–266. New York: Oxford University Press.

Carlson, Eric 1995. *Consequentialism Reconsidered*. Dordrecht Kluwer.

Chang, Ruth. 2015. "Value Incomparability and Incommensurability." In *The Oxford Handbook of Value Theory*, edited by Iwao Hirose and Jonas Olson, 205–224. New York: Oxford University Press.

Dasgupta, Partha. 1993. *An Inquiry into Well-Being and Destitution*. Oxford: Clarendon.

Feldman, Fred. 1986. *Doing the Best We Can*. Dordrecht: D. Reidel Publishing Company.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Hare, Caspar. 2007. “Voices from Another World: Must We Respect the Interests of People Who Do Not, and Will Never, Exist?” *Ethics* 117:498–523.

Holtug, Nils. 2010. *Persons, Interests and Justice*. New York: Oxford University Press.

Holtug, Nils. forthcoming. “Population and Prioritarianism.” In *Oxford Handbook of Population Ethics*, edited by G. Arrhenius, K. Bykvist and T. Campbell. Oxford University Press.

Huemer, Michael. 2008. “In Defense of Repugnance.” *Mind* 117, no. 468: 899–933.

Kavka, Gregory. 1981. “The Paradox of Future Individuals.” *Philosophy & Public Affairs* 11:93–112.

Lazari-Radek, Katarzyna de, and Singer, Peter. 2014. *The Point of View of the Universe*. Oxford: Oxford University Press.

McMahan, Jeff. 1981. “Problems of Population Choice.” *Ethics* 92, no. 1: 96–127.

Mulgan, Tim. 2006. *Future People: A Moderate Consequentialist Account of Our Obligations to Future Generations*. Oxford: Oxford University Press.

Narveson, Jan. 1973. “Moral Problems of Population.” In *Ethics and Population*, edited by Michael D. Bayles, 59–80. Cambridge, MA: Schenkman.

Ng, Yew-Kwang. 1986. “Social Criteria for Evaluating Population Change: An Alternative to the Blackorby-Donaldson Criterion.” *Journal of Public Economics* 29:375–381.

(p. 497) Ng, Yew-Kwang. 1989. “What Should We Do about Future Generations?” *Economics and Philosophy* 5: 235–253.

Ng, Yew-Kwang. 1990. “Welfarism and Utilitarianism: A Rehabilitation.” *Utilitas* 2, no. 2: 171–193.

Parfit, Derek. (1984). 1987. *Reasons and Persons*. Oxford: Oxford University Press.

Parfit, Derek. 2011. *On What Matters: Volume Two.* Oxford: Oxford University Press.

Parsons, Josh. 2002. “Axiological Actualism.” *Australasian Journal of Philosophy* 80, no. 2: 135–147.

Roberts, Melinda A. 2007. “The Nonidentity Fallacy: Harm, Probability and Another Look at Parfit’s Depletion Example.” *Utilitas* 19:267–311.

Roberts, Melinda A. 2009. “The Nonidentity Problem and the Two Envelope Problem.” In *Harming Future Persons*, edited by Melinda A. Roberts and D. Wasserman, 201–228. Dordrecht: Springer.

Roberts, Melinda A. 2011a. “The Asymmetry: A Solution.” *Theoria* 77:333–367.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Roberts, Melinda A. 2011b. “An Asymmetry in the Ethics of Procreation.” *Philosophy Compass* 6, no. 11: 765–776.

Roberts, Melinda A. 2015. “The Neutrality Intuition,” Theoretical Population Ethics Conference, University of Oxford, Oxford (Nov. 2015).

Roberts, Melinda A. 2018. “Does the Worth-Having Existence Make Things Morally Better?” Climate Ethics and Future Generations Kick Off Conference, Institute for Future Studies, Stockholm (Sept. 2018).

Roberts, Melinda A. forthcoming a. “Parfit, Population Ethics and Pareto Plus.” In *Reading Parfit*, edited by A. Sauchelli. Routledge.

Roberts, Melinda A. forthcoming c. “Nonidentity, Better Chance and the Value of Existence: A Defense of Person Based Consequentialism.” In *The Oxford Handbook of Population Ethics*, edited by G. Arrhenius, K. Bykvist and T. Campbell. Oxford University Press.

Roberts, Melinda A. 2019. “Does the Additional Worth-Having Existence Make Things Better?” In *Studies on Climate Ethics and Future Generations* (Vol. I), edited by P. Bowman and Katharina Berndt Rasmussen, 27–40. Stockholm: Institute for Future Studies.

Roberts, Melinda A. forthcoming b. “The Better Chance Puzzle and the Value of Existence.” In *Festschrift for Derek Parfit*, edited by J. McMahan, T. Campbell and K. Ramakrishnan. Oxford University Press.

Roberts, Melinda A., and Wasserman, David T. 2017. “Dividing and Conquering the Nonidentity Problem.” In *Current Controversies in Bioethics*, edited by Matthew Liao and Collin O’Neil, 81–98. New York: Routledge.

Scanlon, Thomas. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Singer, Peter. 2011. *Practical Ethics.* 3rd ed. Cambridge: Cambridge University Press.

Singer, Peter, and de Lazari-Radek, Katarzyna. 2014. *The Point of View of the Universe*. Oxford: Oxford University Press.

Tännsjö, Torbjörn. 2009. “Why We Ought to Accept the Repugnant Conclusion.” *Utilitas* 14, no. 3: 339–359.

Temkin, Larry. 2012. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford: Oxford University Press.

Wasserman, David T., and Roberts, Melinda A. 2017. “Dividing and Conquering the Nonidentity Problem.” In *Current Controversies in Bioethics*, edited by Matthew Liao and Collin O’Neil, 81–98. New York: Routledge.

Weinberg, Rivka. 2016. *The Risk of a Lifetime*. Oxford: Oxford University Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Notes:

([1]) The work started in earnest with a handful of other philosophers four decades or so ago, including most prominently Derek Parfit. It's fair to say that, at the time of his death, Parfit did not regard the problem cases he himself had contributed to the field as fully resolved.

([2]) See, e.g., Arrhenius (2000).

([3]) Larry Temkin makes a similar point. See Temkin (2012, 382–383).

([4]) Among others, cases that give rise to the nonidentity problem, the repugnant conclusion, and the extinction paradox also arise structural issues. See generally Parfit (1987, Part IV).

([5]) This statement of the telic-deontic connection assumes that the case is one in which agents are confident that a given choice will give rise to a given (sort of) future. When the probabilities at stake are less than 1, the connection between the evaluation of the future and the evaluation of the choice is more complicated. See, e.g., Roberts forthcoming c; Roberts forthcoming b).

For purposes here, I follow Feldman in leaving *accessibility* undefined. I'll just note that the appropriate test for accessibility may vary depending on the questions we want to answer. Since here our interest is in describing the relation of *moral* betterness between futures, to say that at a given time one future is *accessible* relative to another is to say that agents (perhaps working together) at that time and in the one future have the resources and the ability to bring about the other future. Thus a future that is *possible* may not be *accessible*; it's *possible* that my new very special puppy can fly, but the future in which he does fly may not be *accessible* relative to this future, that is, the *actual* future. As Feldman proposes, accessibility is best understood as a matter of metaphysics, not epistemology, and should not be confused with *probability*. Thus, on his view, a future may be accessible even if there is nothing agents can do that will make that future particularly probable; the future in which I successfully open the safe's combination lock, without having any clue what the combination is, is accessible even if highly improbable (Feldman 1986, 24–25).

([6]) While we set aside the question of what counts as a *person* for purposes here, as I use it here the term includes many nonhuman animals; surely consciousness, and not species membership, is key. We also set aside the question of what constitutes *well-being*. Perhaps well-being consists in pleasure, happiness, capability, or something else entirely. For purposes here, it's enough to say that well-being is whatever it is that makes life precious to the one who lives.

([7]) See, e.g., Lazari-Radek and Singer (2014, 361–364).

([8]) However, as we shall see, Broome offers an alternate total principle, one that would sum up, not levels of *well-being* (as defined here), but rather levels of the *personal good*. See Broome (2015); see also notes 15 and 16 and accompanying text.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([9]) I borrow this case from Temkin, who uses it to make a different point (Temkin 2012, 135, citing Stuart Rachels). The point I use the case to make here is a point Temkin makes in his *Lollipops for Life* case (Temkin 2012, 34) and Scanlon makes in his *Jones and the TV Transmitter* case (Scanlon 1998, 235 (cited in Temkin 2012, 36–37).

([10]) *Existence comparability* is, that is, an assumption for purposes here. It is assumed that in many cases it's both true and cogent to say that a future in which a given person exists is better, or alternatively worse, for that person than a future in which that person never exists at all. That assumption itself is plausibly grounded in still another assumption: that a person's well-being in any future in which that person never exists at all (nonexistence itself being devoid of all plusses and all minuses; of all pleasure, all pain, all happiness, all unhappiness, all capability, all incapacity, all of that which makes life precious, all of that which can make life *less* than worth living) is just zero. Existence comparability is a matter of ongoing controversy (Bader 2019; Holtug 2019; Roberts forthcoming c; Holtug 2010).

([11]) Others, including Hare, have used the term *moral actualism* (Hare 2007). Hare describes a second form of moral actualism as well, what Hare calls *weak* moral actualism (Hare 2007, 502–503). On that view, the people morality concerns itself with are those who do or will exist under the choice under scrutiny. That principle quickly leads to inconsistency, and I set it aside for purposes here. See Hare (2007) and Roberts (2011a, 2011b).

([12]) Narveson (1973, 73).

([13]) Weinberg (2016); Parsons (2002).

([14]) Here, we again assume existence comparability. See note 11.

([15]) Broome (2015).

([16]) I have previously proposed such an extension of Broome's concept of the personal good (Roberts 2015; Roberts 2018; Roberts 2019).

([17]) And moral actualism is a view we shall want to reject. See section 4.3.

([18]) The proposal here is captured in the *Loss Distinction Thesis* (Roberts forthcoming c; Roberts 2019), a principle I previously called *variabilism* (Roberts 2011a; 2011b). According to that thesis, a well-being *loss* sustained by p at x relative to y is *morally significant* if and only if p does or will exist at x. Moreover, a well-being *gain* accrued by p at x relative to y is morally significant if and only if it reverses a morally significant loss. For the argument that a broader principle, one that would deem a *gain* significant if the person does or will exist in the future in which the *gain* is accrued fails, see Roberts (2011a; 2011b).

([19]) McMahan (1981). See also Singer (2011, 114–119); Roberts (2011a, 2011b).

([20]) Again we assume existence comparability. See note 11 (on existence comparability).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([21]) Parfit (1987, 363).

([22]) See section 4.1.3.

([23]) Does Parfit himself mean to evaluate the choices as well as the futures? Often he has said that he doesn't. He nonetheless writes in connection with, e.g., his example of Hell Three that "[o]n the average principle, we *ought* to have these children … whose lives would be much worse than nothing" (Parfit 1987, 422, emphasis added).

Another question is whether the populations in A, A+, and B are meant to overlap at least in part. Graph 25.2 assumes that they do, an assumption supported by Parfit's account of the notion of the *mere* addition, which obtains, he writes, only in the case where the additional people "affect no one else" (Parfit 1987, 420). For Parfit's original presentation of the case, see Parfit (1987, 419–420).

([24]) This particular statement—there are many—is based on Parfit's own first articulation of the paradox. See Parfit (1987, 425–430).

([25]) Parfit (1987, 419 and 426).

([26]) Parfit (1987, 381–390).

([27]) Parfit (1987, 388).

([28]) Parfit (1987, 420).

([29]) Parfit (1987, 426).

([30]) Parfit (1987, 426).

([31]) *Pareto plus* is Partha Dasgupta's name for the relevant principle (Dasgupta 1993, 382–383).

([32]) Indeed, let's just note the Ng, for example, considers the rejection of Pareto plus, at least in the context of many cases, "illogical" (Ng 1989, 241). For discussion of Ng's position, see section 4.2.2.

([33]) I have elsewhere presented the argument that MAP, given certain widely held assumptions, implies Pareto plus. See, e.g., Roberts forthcoming a; Roberts forthcoming b; Roberts 2018.

([34]) Depending on whether we can extend Broome's concept of the personal good to reflect our existential values, however, the alternate form of totalism that Broome describes may be consistent with the claim that MAP is false. See notes 15 and 16 and accompanying text.

The formulation of the traditional total view—a view that references raw, unadjusted well-being, *not* an extended concept of the personal good—that I outline here is just one among many extensionally equivalent formulations of the traditional total view.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([35]) Parfit (1987, 422). Parfit also notes the "Egyptology" objection as a basis for the rejection of averagism. See Parfit (1987, 420) ("research in Egyptology cannot be relevant to our decision whether to have children").

([36]) Such an approach, however, isn't necessarily inconsistent with Broome's alternate formulation of totalism. See Broome (2015); see also notes 15 and 16 and accompanying text.

([37]) I should note that most philosophers construct the person-based intuition in just that way. My own view, however, is that that construction dooms the person-based view to inconsistency. Thus for purposes here I reject a restrictive PBI in favor of an expansive PBI.

([38]) I have elsewhere argued that expansive PBI is consistent with the principle of the independence of irrelevant alternatives. My argument relies on the *accessibility* relation—here, a relation based on what agents have the resources and ability to bring about; but a relation that can be constructed in other ways depending on the task at hand—and claims that, where a future y is *accessible* relative to a future x, it's *necessary* that y is accessible relative to x. If we say, e.g., that agents in x have the resources and the ability to have brought about y *and only y* in place of x—and thus that y *and only y* is *accessible* relative to x—we can't then take the position that we can add a third accessible future z to the case without moving from a comparison of x and y to a comparison of some x′ distinct from x and some y′ distinct from y (Roberts forthcoming c; 2018b; Roberts 2019).

([39]) Parfit briefly considers, and rejects, the idea that we might cite B as part of a person-based account of why A+ is worse than A. See Parfit (1987, 428–429). Some of his discussion at this point suggests that he means to invoke the principle of the independence of irrelevant alternatives to argue that whether A+ is worse than A *cannot* depend on whether B exists as an alternate accessible outcome. See note 38.

([40]) See generally Parfit (1987, 351–379) and Kavka (1981). See also Roberts (forthcoming c).

([41]) Parfit (1987, 361).

([42]) The term is Kavka's, who described the "precariousness of our origins" (Kavka 1981, 93).

([43]) See Roberts (2007; 2009) Roberts and Wasserman (2017); Roberts forthcoming c.

([44]) Parfit (1987, 387–391).

([45]) Broome's formulation of totalism, according to which well-being levels above the zero level but below a certain critical level count against a given future, has the resources to avoid some versions of the repugnant conclusion (Broome 2004). Arguably, however, it won't avoid them all (unless we stipulate a very high critical level, a strategy that will force us to say that even the mere addition that is *maximizing* for the additional person

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

makes things worse and is wrong). Moreover, the critical level form of totalism appears to be vulnerable to what Arrhenius calls the *sadistic conclusion* (Arrhenius 2000).

([46]) Huemer thus offers a defense of the repugnant conclusion. See Huemer (2008). Tännsjö as well urges us not to be too quick to reject the repugnant conclusion, on the grounds that, even if initially repugnant, the conclusion is not "*obviously* wrong" and, moreover, is an implication of "*any* plausible moral principle" (Tännsjö 2009, 339).

([47]) Ng (1989, 237). Thus, according to Ng, without Pareto plus, we are unable to resolve what he calls the *extinction paradox*. Philosophers—including Ng—seem to have very strong feelings on the question of the disvalue of an earlier, rather than a later, extinction of the human (or any successor) species. One philosopher at a recent conference described early extinction as the worse "evil" he could imagine, even in the case in which the extinction event itself (somehow) made things worse for no one, including members of the last generation. But intuitions vary. Others may think that it would make things worse, and be wrong, to make things worse for existing and future people just so that the species itself could be perpetuated more or less indefinitely.

Pareto plus is also, according to Ng, a prerequisite for a plausible solution to the nonidentity problem. Thus for Ng what makes it clear that the choice of depletion makes things worse and would be wrong is that the *merely possible* people who never in fact exist under depletion but who would have existed under the choice of conservation themselves have a sort of claim or right to be brought into existence that ought to have been weighed against the parallel claim of the people who in fact exist (the people who exist, that is, and are burdened, by the choice of depletion).

Ng's thinking here seems to be that, since the people who would have existed under conservation have more to lose—and more to gain—than the people who do exist under depletion, the people who would have existed under conservation have the superior claim. But even if we assume that all those people *do* have the claim that Ng imagines, it's unclear why we should think that the claim of the potentially better-off group is superior. As noted earlier, however, we in any case arguably have options for addressing the nonidentity problem that do not depend on Pareto plus. See section 4.1.3.

([48]) Ng (1989, 237).

([49]) Ng (1989, 241, citing Simon).

([50]) Ng (1989, 237).

([51]) Ng (1989, 241).

([52]) See Carlson (1995, p. 100).

([53]) Moral actualism raises still other difficulties as well. Consider, e.g., Parfit's Tom, Dick, and Harry case. Parfit discusses variations on that case in Parfit (2011, 223–331). There we want to say that the three futures are each exactly as good as each other. Moral actu-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

alism, however, rules out that evaluation. Also see also Hare (2007, 504, Jack and Jane case; and 509, Jack, Jane, and Fred case).

([54]) Temkin (2012, 368–380).

([55]) Temkin (2012, 371–376; see generally chaps. 6 and 7). Importantly, the varying relevance that Temkin describes in this case isn't subjective in nature, a matter of the values *we take to be* relevant to the comparison. Rather, the relevant values may be an objective matter of fact, determined, it seems, by the relationship between the two futures that are the subject of the comparison (Temkin 2012, 373).

([56]) See, e.g., Temkin's tennis players and husbands example (Temkin 2012, 380). One man might be a better husband than another and that second man might be a better tennis player than a third. But it won't follow that the first man is a better husband or a better tennis player than the third.

([57]) Among the most promising of our further options are Mulgan's rule utilitarianism (Mulgan 2006, 55–81) and Holtug's wide person-affecting approach (Holtug 2010).

Melinda A. Roberts

Melinda A. Roberts is Professor of Philosophy at the College of New Jersey and recently completed a Laurance S. Rockefeller faculty fellowship at the Princeton University Center for Human Values. Both a philosopher and a lawyer, she is the author of Child Versus Childmaker, Abortion and the Moral Significance of Merely Possible Persons and a number of articles in the areas of population ethics (including the repugnant conclusion and the nonidentity problem), procreative ethics (including wrongful life and reproductive technologies), and climate ethics. She continues to have an interest in developing a person-based form of consequentialism that functions well for both the evaluation of choices and of outcomes.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Conflicts and Cooperation in Act Consequentialism

Joseph Mendola

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.21

Abstract and Keywords

There are circumstances—involving conjunctive acts at a time or over time or the conjunction of acts of different agents—that threaten act consequentialism with self-defeat. There are also variants of act consequentialism—including consequentialist generalization, generalized act consequentialism, multiple-act consequentialism, cooperative consequentialism, and modally robust act consequentialism—that promise a more plausible treatment of circumstances in which cooperation is important than traditional act consequentialism. Links among these issues are explored. Attention to the nature of the normative conflicts occasioned by consequentialist assessment of actions can help reveal an important and useful variety of forms of, and resources for, act consequentialism. Its structure is not as simple as it may initially seem.

Keywords: self-defeat, the principle of moral harmony, act consequentialism, cooperation, group act, consequentialist generalization, generalized act consequentialism, multiple-act consequentialism, cooperative consequentialism

CONSEQUENTIALISM assesses alternatives by regards to the value of their consequences. And so it is not surprising that consequentialist assessment of different types of alternatives—say of acts, virtues, decision procedures, rules, and practices of punishment and blame—can generate normative conflicts. What it is best to do on a very unusual occasion may sometimes be best punished as a general practice; a person with the best character may sometimes fail to decide in a strictly optimific way; the best possible moral rules for a society may sometimes require action it is not best to perform, for instance when those rules are not generally followed by others. What is more surprising is that there can be normative conflicts generated by consequentialist assessment of a single type of alternative, for instance by consequentialist assessment of actions, as we will shortly see. But there is a benefit hidden in this unfortunate surprise. Attention to the nature of the normative conflicts occasioned by consequentialist assessment of actions can help reveal an important and useful variety of forms of, and resources for, act consequentialism. Its structure is not as simple as it may initially seem.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Act consequentialism assesses acts by the value of their consequences, relative to that of the value of the consequences of alternative actions open to the agent. In a straightforward case, where a person at a time only has three alternatives—act A, act B, and doing nothing—and the consequences of A will be better than those of B and of doing nothing, it stipulates that the person do A. It specifies that A is what ought to be done, or what is right.

There are, however, important and significant complexities that such a simple case neglects. In reality, one often has an alternative that involves various probabilities of various outcomes and not just the certainty of one outcome. In reality, one often is ignorant or epistemically uncertain about what one's alternatives are. Arguably, there are sometimes objective indeterminacies in one's alternatives, even when all the facts are known.[1] (p. 514) And, on some views, unlike classical utilitarianism, the value of outcomes is relativized to agents, so that what is a better outcome from the perspective of one agent is not a better outcome from the perspective of another. But the sorts of conflicts we need to consider are most clearly revealed if we idealize these complexities away, if we stipulate that there are in a given case a specific set of alternatives open to the agent, each of which involves determinate and certain outcomes that are fully known, with specific objective values that do not differ from the perspective of different agents. This will allow us to focus on other, more relevant complexities.

1. Conjunction Problems and Self-Defeat

The first type of normative conflict within act consequentialism that we will consider is due to the fact that some acts are conjunctions of other acts.[2] There are several subtypes, but let's start with an easy case.

At some particular time, you may be able to make a certain aggressive hand gesture or refrain from it, and you may also be able to concurrently smile broadly or maintain your cool and neutral expression. If you make the hand gesture while smiling, all will be well, better than if you do nothing at all. But if you make the hand gesture while keeping your flat expression, the social effects will be toxic. So too if you simply smile. It is optimific to make the gesture and to smile, but it is not optimific to gesture, and not optimific to smile. And yet to gesture and smile is just the conjunction of gesturing and smiling. It seems that act consequentialism both recommends those actions and forbids them.

What is the act consequentialist to do? In such a case, it is not hard to see. The proper conception of your alternatives at the time focuses on maximal alternatives, alternatives to do every combination of intuitive acts (and nothing) you might do at the time. There are in this case four relevant maximal alternatives: You can do nothing, you can smile without gesturing, you can gesture without smiling, and you can gesture and smile at once. The fourth is what act consequentialism requires. And the second and third alternatives are worse than the first.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

But more interesting conjunction problems involve time.[3] Some intuitive acts of a given individual are conjunctions of more momentary actions.[4] Baking a cake involves several sequential steps. You can take a certain step in the process and then you can take another. If you take both steps, things will go very well. But if you take just the first step alone, or just the second alone, it will be a wasteful disaster. Perhaps the relevant alternatives seem here again to be four: Take both two steps. Take just the first and skip the second. Skip the first and take just the second. Or do nothing. Those are temporally (p. 515) extended alternatives that you could in some sense take, and the first alternative is consequentially best. That is analogous to our solution to the problem of conflicts at a time. But this case is more controversial. Notice that it is not obviously open to you right at this moment to assure that if you take the first step you will then subsequently take the second. You may have well-founded worries about your character, and realize that, easily distracted and weak-willed while baking as you are, you won't in fact take the second step when the time comes, even though you have already taken the first, and even though in some sense you could take the second. Act consequentialism may seem to give contradictory advice in such cases. It may seem at once to require that you take the first step as part of a longer project open to you, and also to forbid that first step given your character. And that seems inconsistent.

Still, this dispute, like the first dispute involving a putative conflict at a single time, can be assimilated to the issue of what in fact the proper conception of the relevant alternatives is, although in this case there are different and competing views about that which are popular among philosophers. Perhaps your relevant options right now are constrained by what in fact you will do in the future, or at least by what you can now assure that you will do in the future. And on such conceptions of your current options, act consequentialism recommends not taking the first step.[5] On an alternative conception of your options, focusing on what is up to you at each point going forward, act consequentialism recommends taking the first step as part of a longer project. One way to put this issue is that philosophers who disagree about the relevant conception of alternatives in such cases differ about what one is properly held morally responsible for, whether one's foreseen character flaws are just part of the context in which one now acts, or rather more intrinsically relevant to the moral character of the act one is now performing, whether one can be held responsible now to perform future acts that one cannot now secure. So at least these normative conflicts within act consequentialism can be potentially resolved by attention to issues about alternatives and responsibility. Any complete and consistent moral theory in the consequentialist tradition will need to resolve these issues in one way or the other, because it will need some consistent account of alternatives and responsibility. We will return to this matter eventually. But there is yet a third cluster of analogous conjunction problems that cannot be so easily assimilated to issues about alternatives and individual responsibility.

There can be two agents so situated that if, at the same time, they each do their part in some cooperative scheme, the consequences will be very good, but such that if one plays their role in that scheme while the other neglects their role, the results will be disastrous. This raises a number of questions for the act consequentialist. Should the consequences

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

of the action of the first be assessed in the context of what in fact the other will do at the same time? Or, to relax our earlier idealizations, is what the other will probably do most relevant? Or is what is relevant whether the first cooperator can assure that the second will do their part in the scheme, if indeed anyone can ever assure what someone else will (p. 516) do in that way? Alternatively, if the two agents are cooperating in some way, or should cooperate, should we focus moral evaluation most centrally not on what they do as individuals but on what they can together accomplish? There are a variety of normative puzzles and conflicts that arise for act consequentialism out of such cases, and out of more complex cases involving many agents acting over time. But we will focus on just a few examples that are instructive in other ways.

Act consequentialism directs individuals to act for the best. So consider a situation in which a number of individuals individually act for the best, as act consequentialism directs. Might it be that what they together accomplish is not as good as what might be accomplished if they individually acted in other ways? If so, that would generate one kind of normative conflict within act consequentialism. It would mean that act consequentialism violates what is sometimes called "The Principle of Moral Harmony," that if everyone does their moral duty according to a moral theory in their given circumstances, then the world will be as morally ideal as it can be according to that theory in those circumstances.[6] Since act consequentialism is focused normatively on the production of good consequences, it would mean that there is one kind of conflict within act consequentialism itself.

However, there may seem to be a successful argument that such a conflict cannot arise for act consequentialism, at least under the conditions of certainty, full knowledge, and determinacy of objectively valuable alternatives that we presume: If any given individual from among the group in question has a choice among individual options of which one option is optimific, and they fail to take that optimific option, then the overall consequence of what all the individuals acting together will achieve will hence be worse than it would otherwise be. But if all of the individuals take their optimific options, then each will have made their best available contribution to overall consequences. So it seems that if each individual agent acts as best they can, then the group will do so as well.

It sounds plausible. But there are significant gaps in this argument.

Consider a situation in which Eve and Adam are in the following circumstance, redolent of the Prisoner's Dilemma:[7] Each has two options, to push a button or not. If they both push, then the outcome is an excellent 10. If they both don't push, then the outcome is a middling 5. But if one pushes and the other doesn't, if they fail in that way to cooperate, the outcome is a lousy 0. For each agent, which option is best depends on what the other party does. Each does the best they can, on the presumption that the other doesn't push, by not pushing. And in that circumstance, while they both do what act consequentialism requires, they do not act together to assure the best outcome available to them. The Principle of Moral Harmony is violated.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

What went wrong in our plausible argument? Two things. The first thing that went wrong is that there are two outcomes that can be achieved consistent with our pair acting as act consequentialism demands of individuals in this situation, but only one is optimific overall, only one is such they do as well as they can acting together. The optimific outcome is achieved when each does individually the best they can in the specific (p. 517) circumstance of the other pushing, by pushing themselves.[8] But each also does the best they individually can in the other specific circumstance, where the other doesn't push, by not pushing.

Still, this case doesn't show that *all* ways for individuals to follow act consequentialist strictures in such a situation will entail a loss of good consequences for the group acting together. There are two patterns of individual act consequentialist activity in question in the case, and one is optimific overall. It may seem that there must always be *some* way in which such individuals do what act consequentialism demands of them and yet the overall achievement of the group is optimific.[9]

However, there is a second significant gap in our apparently plausible argument. That argument presumes that the relevant natures of the alternatives available to agents do not depend on the kinds of choices among alternatives the agents make. And this can be false. Powerful and playfully evil Martians may announce that if each individual in some group of human agents chooses among their options in the way that act consequentialism demands, chooses the consequentially best option available to them, then they will make all the options of all the human agents worse in consequential terms than they otherwise would be, say by torturing all humanity for a millennium.[10] The Martians may even announce that only if *all* of the agents do less than their individual best, will they allow humanity to escape completely unscathed. The group would do better in consequential terms if all the individuals in the group acted in other than an act consequentialist way. How can this be? Because there is a kind of modal instability in this situation. If one of the agents chooses the second best option B of some set of alternatives open to them, it is true that they had a better option A available to them which they did not take.[11] But had they in fact chosen that option A, it would have hence been much worse, much worse than B will be if the agent takes option B.[12] The same sort of argument can reveal an analogous sort of conflict within individual act consequentialism itself. The troublesome Martians can in the same way make all the options of any given individual act consequentialist much worse than they otherwise would be, even if no group of agents is involved.

(p. 518) So there are possible circumstances in which act consequentialism violates the Principle of Moral Harmony. It can be self-defeating in that way. But of course the Martians are wildly hypothetical. Maybe such wildly hypothetical circumstances are irrelevant to ethics. Still, we know that, here in reality, even ignoring the ignorance and uncertainty ruled out by our idealizations, act consequentialism has costs in its own terms.[13] For instance, the many who are not act consequentialists may find it hard to trust act consequentialists to tell the truth and keep their promises, since they will only do so when it is optimific. They are even prepared to murder cooperators should utility be so served, and that may be a significant barrier to trust! And so many of us may find it hard or im-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

possible to cooperate with act consequentialists in consequentially weighty cooperative schemes. And it is certainly not obvious that the world would be the best place we could together make it even if all acted as act consequentialism demands. It is not even obvious that the world would be the best place we could together make it if all acted in *some* way that individual act consequentialism demands. For all we have seen, act consequentialism may sometimes be self-defeating in reality in the rough manner suggested by the Martian case. Still, even if it is, there would be the further question of how common and important these conflicts are, or rather how common and important they would be if act consequentialism were widely followed. Another question would be whether they show anything of normative significance. I think that the Martian case is sufficient to show that self-defeat per se isn't very normatively significant. Any normative theory that grants significant weight to people's well-being, which any good normative theory should, can suffer self-defeat in such a way in at least hypothetical circumstances, and that shows such self-defeat is not very important. Still, the other case we have considered, the Eve and Adam case, by focusing on cooperative action, which is, unlike the machinations of the Martians, central to morality here in the real world, may point us toward a serious difficulty for act consequentialism. While the Eve and Adam case itself doesn't show anything quite so strong, still if act consequentialists cannot adequately cooperate, cannot adequately work together in pursuit of the good, then act consequentialism may be in serious and revealing tension with its own goal and rationale. That would be a kind of self-defeat that matters.

2. Cooperation

Can act consequentialists cooperate? How can they be trusted to take part in an agreed cooperative scheme if, at any moment that better consequences become available to them, they will defect from the cooperating scheme and grab those extra good consequences on the side?

There is some good news on this topic: If the cooperative scheme is optimific, and there is a history of ongoing cooperation between our consequentialists, and various (p. 519) other reasonable conditions are met, it is provable that they will continue to cooperate.[14] However, not all our actual forms of consequentially weighty cooperation are strictly optimific. Real life is not that ideal. And we might not have a settled pattern of cooperation to depend on.

Unfortunately, there are realistic circumstances in which act consequentialists cannot do what best cooperative action requires, even if we suspend worries about trust. There is a boulder perched on a hill above a village of act consequentialists, which is going to topple down and destroy the village. Together they could push the boulder, so it would fall in a harmless direction. But the greatest good they can each assure acting on their own is the survival of their own family and possessions, helping their neighbors do the same when it is optimific to do so. They each reason that they can't move the boulder on their own, and so they shouldn't be deflected into trying, because that would hobble their attempts to save their own family and neighbors, when they know that because of similar reasoning

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

the other villagers will make the same decision to save their own.[15] Because of urgency and the lack of relevant social arrangements, the villagers cannot do anything consistent with their act consequentialism to escape this unfortunate circumstance. If only their moral principles had instead designated some village elder the authoritative decider, or otherwise specified the right form of cooperation for such circumstances!

This is at least potentially a serious problem for act consequentialism. It seems to reveal a kind of important self-defeat, in morally significant conditions. There are other hoary traditions in ethics that focus on the significance of what we do together, or of what would be true if we all did something even though we don't all do it. Maybe these traditions recognize something important about morality that act consequentialism slights.

But act consequentialists shouldn't give up too easily. Perhaps there are ways for act consequentialists to embrace these insights of other traditions. Perhaps act consequentialism can be modified in some way consistent with its optimific spirit to support optimific cooperative activity. Or perhaps it is a mistake to think that act consequentialism itself properly focuses only on individual acts. There are various mechanisms or modifications that might link traditional act consequentialism and beneficent cooperative activity. So let's turn to a consideration of that range.

One such mechanism is *consequentialist generalization*. Perhaps to choose an individual act from among individual options is at least implicitly to choose an act type to be performed by all others in relevantly similar circumstances.[16] On this conception, the relevant options to some act one performs encompass not just one's individually available alternatives but also the social alternatives to others acting in a similar way in similar circumstances. And it is the consequences of everyone acting in such a way that (p. 520) matter. Accepting consequentialist generalization would for instance force both Eve and Adam, and all our villagers, to the optimific alternatives for the relevant groups, and that may seem a good thing.

But one serious problem with this idea is it can sometimes be a disaster to do your part in some cooperative scheme when someone else does not do their part. If all the villagers but one stick to their individual act consequentialist activity of saving their own, while one lone moralist does what consequentialist generalization requires, their family will be killed and nothing will be gained. Consequentialist generalization does not properly distinguish between circumstances in which someone should do their part in a cooperative scheme and others in which they should not because not enough others are doing their part. This is also a serious problem for *rule consequentialism*, which by focusing on acts required by the consequentially ideal moral rules for a society to accept or follow, on such an indirect consequential assessment of acts, is in any case outside of the act consequentialist framework we are now exploring.[17]

There is also another problem case for consequentialist generalization:[18] Someone has fallen through the ice while skating, and there are twelve act consequentialists arranged symmetrically around the pond. If all the consequentialists rush to the rescue, the ice will collapse and the outcome will be horrific. So it would be better, on the presumption that

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

not all will act, that one of the individuals go to the rescue, in contradiction to consequentialist generalization. In general, some cooperative schemes depend for their success on different individuals doing different things, performing different roles, and maybe doing nothing at all. And consequentialist generalization is not well-equipped to deliver such success.

So let's try another mechanism, *generalized act consequentialism*.[19] There are, at least arguably, things we together do, genuine group acts. So maybe we should assess the situation by the pond as act consequentialists, but from the perspective of group actions. The group by the pond should distribute responsibilities and act together in the optimific way, more or less analogous to the way an individual might work over time in the most effective way available to that individual by doing different things at different times. If the group doesn't work together in the right way to save the person who has fallen in, then the group has not acted in the proper consequentialist way, not even the proper act consequentialist way, we might claim. That is because among acts are group acts.

Still, there are a variety of different group acts that might be performed by any group of people, while there may be nothing a given individual person can do to assure that some group will act together in a certain way rather than another. There won't be time for the act consequentialists around the pond to appoint a committee to decide on their individual roles to save the drowning skater, nor for the villagers to have a town meeting. So responsibility does not seem to distribute over groups in the same way as it distrib (p. 521) utes within an individual over time. And while generalized act consequentialism can condemn a failure of cooperative activity in which the villagers merely save their own, still there would be nothing that the individuals involved in such a circumstance would be doing wrong as individuals according to that view, despite the failure of group action. The group act is condemned, but not the individual acts which make it up, which itself is a kind of normative conflict within act consequentialism.

Our act consequentialists around the pond and in the village have to come up with group acts out of nothing, and since there are many possible group acts they might perform, some of which are very beneficent but not strictly optimific, it may be hard for them to settle on a single cooperative scheme, a specific group act. There is a relevant difference between these two cases. One thing that may assist our act consequentialists around the pond is that they can see how many of the others are in fact rushing to the rescue, and then act individually on that basis. But the act consequentialists in the village face especially problematic circumstances for act consequentialism. They cannot as individuals achieve their best good unless many others in symmetrical positions positively cooperate very quickly in exactly the right way.

In general, if group acts are going to be an important resource for act consequentialism, we need some specificity of relevant group action, to fix the proper roles of individuals in the proper group act, which in turn will help to determine when the failure of an individual to follow their role is morally wrong. Such specificity can be contributed in some cases by actual group acts in which specific individuals already participate, so that we can

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

see what their individual roles are in that group action.[20] There are a variety of different accounts of what such group acts are, and while some think the notion just a metaphor, it is not implausible to claim that sometimes a group of individuals literally do something together, that a department pursues a goal or a couple tries to work it out. And on many such accounts, group acts can specify distinct particular roles for different individuals. Someone is supposed to provide distraction while the other party to the scheme goes in for the kill.[21]

There are a variety of advantages that accrue to a focus on actual group acts in ethics.[22] For instance, our deontic moral practices, such as that forbidding lying at least under many conditions, or that supporting intuitive practices of distributing harms in trolley cases, may be consequentially weighty group acts. And so a focus on group acts may evade some traditional objections to act consequentialism. It still may seem a puzzle why it would be appropriate for an act consequentialist to continue to participate in an actually beneficent group act, to do their part, when they can gain a little extra positive consequence on the side and yet the group act not collapse, yet achieve its beneficent goal. Many of the actual group acts in which we participate create great good but are not strictly optimific. They involve surplus cooperation, in the sense that the occasional (p. 522) defector can gain something good while there is no compensating loss in achieving the goal of the group act. Perhaps someone could tell just one white lie to grab some good on the side while the general group act forbidding lying sails on unperturbed. However, the intuitive advantages of a focus on actual group acts often depends on surplus cooperation of exactly this sort, so that for instance our deontic practice forbids the occasionally beneficent murder of the innocent. What's more, intuitive exceptions, truly white lies, would often be at least implicitly accepted by participants in the relevant group act, and so plausibly built into the content of the group act. When defections would be criticized by other participants in the group act, it seems that surplus cooperation is supported by the same sorts of intuitive motives that support contractualist conceptions of ethics rooted in reciprocity, while the beneficence of the group act itself adequately respects intuitive motives like benevolence that support consequentialist conceptions. In addition, such defections are analogous to individual momentary defections from continuing individual plans of best action over time, and might be thought inappropriate or irrational on similar grounds. So there is a good deal that can be said in support of surplus cooperation.

Still, this conflict between the good that an individual may accomplish on their own and what a beneficent group requires of them is one type of normative conflict that an extension to actual group acts can introduce into act consequentialism. And actual highly beneficent group actions that are not strictly optimific may be very important in production of the good, while yet not better than all available alternative group acts. And so pursuit of the best of any set of alternatives, group or individual, may require defection from group acts with very good consequences. What's more, there is yet another type of conflict introduced into consequentialism by a focus on actual group acts: An individual person may be at once a participant in two beneficent group acts, say involving different

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

groups of cooperators, that specify conflicting roles for that individual. They can't do their part in both. What should they do?

One way to handle such conflicts is by the mechanism of *multiple-act consequentialism*:[23] This requires that when different beneficent group acts of which one is part specify roles that conflict, one ought to follow the role in the group act with more valuable consequences. One should defect to the dominant group act. Why? Because it is the act with the weightier consequentialist rationale. Especially since individual acts over time are a kind of group act that one performs with oneself, in which different periods of one's life cooperate in an overall goal, it is natural to extend this treatment to another case: One should only defect from a group act with good consequences if one can achieve better consequences by the defecting act alone than the entire group act achieves. Very little defection is allowed. When should one join a group act? When it is consequentially best to do so. This means there is an asymmetry in the conditions that must be met for appro (p. 523) priately joining and defecting from group acts, which by a kind of ratchet draws us into beneficent group acts from which we can no longer properly defect.

I have argued elsewhere that multiple-act consequentialism is quite a magic bullet in answering traditional intuitive objections to act consequentialism, since it delivers traditional deontic duties and virtues in the most forcefully intuitive cases.[24] But there is a serious limitation of this mechanism. By its focus on actual group acts, rather than on cooperative schemes we might engage in, it still apparently fails to deliver our genuine moral obligations not only in hypothetical cases like those we have been considering, but also in some very important cases that we face here in reality, real-world cases that threaten important sorts of conflict or self-defeat for act consequentialism.

One of the greatest moral challenges facing us today is radical climate change due to human activity. But individuals are not, at least in any obvious way, participants in actual group acts that specify roles that could deliver us from that challenge, nor does traditional act consequentialism seem to provide an adequate response. We have no widespread deontic practice that might be considered a weightily significant group act and which forbids doing all the little things we do to make the climate change for the worse. And while it may seem that the moral significance of consequences requires urgent action by all of us together to limit the effects of climate change, any individual consequentialist acting on their own can do little to halt its inexorable progress. Indeed, a given act consequentialist should perhaps not even refrain from their individual contribution to worldwide pollution, since it may be in the service of some small local good, and no individual can do anything on their own about the grand climatic tragedy upon us. If this is so, we seem to be stuck in a situation very like the villagers threatened by the boulder. Act consequentialism may direct us all only to worsen our most consequentially weighty normative problem, rather than work together to somehow secure better consequences for all. That would be a rather grand and terrible form of self-defeat for act consequentialism.

There are attempts to answer this worry within the strictures of act consequentialism, at least if we relax some of the idealizations I have made. Perhaps there is some chance,

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

however small, that one's individual pollution will be a tipping point that leads to catastrophe, and the small risk of a horrendous outcome may be a consequentially terrible risk.[25] But still, there are some realistic situations in which attention to risk cannot dispose of the worry. Two individual act consequentialists may own factories which are such that the continued operation of each factory alone is sufficient to destroy some significant natural wonder, but such that each does some good in supporting workers in a local community. This is a case of surplus harm. Each consequentialist can reason that the natural wonder will be destroyed whatever they do, and yet the small good they do will be achieved if they keep on polluting. So they may seem locked into the poor outcome by act consequentialism.[26] Our circumstances regarding climate change may be quite analogous.

(p. 524) Act consequentialism needs some other response. Perhaps we can get to the right place by arguing from the inside out, given our actual group acts. Perhaps there is a large group act of selfishness in which we generally engage and which has horrific consequences, so that we must defect from that group act to be minimally moral. And perhaps the only way to defect implies ceasing to pollute at all.[27] But whether or not this captures the evil of some of our actual polluting activity, still there are realistic kinds of pollution, aiming always locally to maximize the available good, in the manner of our two factory owners, which cannot be condemned by this mechanism. Our two factory owners, unlike many real ones, aren't selfish.

So we need another response. Perhaps some of our actual group acts harbor potencies that will be sufficient to solve this problem. Perhaps citizens participate in a group act that is their system of law and government,[28] and perhaps our polities will work themselves toward adequate laws of the requisite sort. But while we can hope so, it is not wise to count on it.

3. Novel Cooperation

Some other elaboration of act consequentialism seems necessary if it is not to be condemned in its own terms. It seems that we must focus somehow on the creation of specific novel group acts or cooperative schemes.

There are mechanisms that have been proposed to do so. *Cooperative consequentialism* specifies that each individual agent ought to cooperate, with whomever else will cooperate, in the production of the best consequences possible, given the behavior of the noncooperators.[29] An element of actuality is contributed, according to this conception, by who is prepared to cooperate, but then the nature of the group act in question is up for grabs, up for proper consequentialist determination.

And there is also a second mechanism that has been proposed to stir a focus on novel cooperation into act consequentialism. To understand it, consider first this variant of the case of our factory owners: One owner is prepared to cooperate with the other and close the factory if only the other will, but the second is not prepared to cooperate in that way.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

While apparently both in actuality do what act consequentialism requires, like the original factory owners, when they keep their factories open, in this case there is an asymmetry. The first owner has a virtue of cooperativeness in pursuit of the good that the second lacks, because the first would do what act consequentialism demands in a hypothetical situation, in which the other cooperates in the proper way, but the second would not. It may seem that traditional act consequentialism cannot deliver the evident moral difference between these two owners. And the need for a moral difference between them may suggest another mechanism for cooperation. *Modally robust act* *(p. 525)* *consequentialism* specifies not only that an agent ought to act optimally in the actual world, but be such that for all possible combinations of the actions of other agents, if that combination were instantiated, he or she would act optimally.[30] In our recent case, the second owner would not act optimally if the first did the cooperative thing, but the first would if the second did. So the first satisfies the requirements of modally robust act consequentialism while the second does not. And this delivers the intuitive moral asymmetry we want.

But notice that the second owner really doesn't even in actuality do what act consequentialism demands in this situation. That is because the first owner has the virtue of cooperativeness, and so if the second owner had acted cooperatively, then the first would have as well. So there is a problem with this case. If the purpose of this modification of act consequentialism is only to find grounds to morally condemn one of the two owners, it is important that this condemnation is already available to the standard act consequentialist. Still, like cooperative consequentialism, this proposal also points in the cooperative direction we are now exploring, and serves that other purpose. Nevertheless, to the degree that this proposal is different from cooperative consequentialism, it plausibly has consequential costs. The only truly beneficial cooperation that occurs will exist in reality. And this even cooperative consequentialism assures. But the robust conditions of character required in addition to that by modally robust act consequentialism, required for it to be true of someone that they would cooperate beneficently in wildly hypothetical circumstances, may well be consequentially harmful in reality while not in reality generating any extra good. So cooperative consequentialism seems to be the better mechanism of the sort now under consideration.[31]

Some problems remain. Cooperative consequentialism, like modally robust act consequentialism, apparently requires cooperators who are not only able but willing to cooperate in pursuit of fully optimific alternatives, and such ambitious and beneficent cooperators may be in short supply in many important real cases. Perhaps the individual sacrifices required by strictly optimific group action to halt climate change are very great, while much less sacrifice would be required for adequately responsive group action, at least if we include in the relevant group act not just those who might conceivably be cajoled into the great sacrifice required by strictly optimific group activity, but also the greater number who would be willing to do their part in something less terrifically demanding. And it isn't always obvious what mechanism for deciding on a single specific group act is available to the cooperators. So perhaps we should introduce some modifications into cooperative consequentialism.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Here's one way to do that: Perhaps the proper form of act consequentialism should encourage novel cooperation not only with those who are prepared to act cooperatively (p. 526) for the best, but with those who, while not prepared to do whatever is required to do their part in some very demanding and strictly optimific group act, are prepared to work in some positive way, to join at least many new beneficent group acts that will help. And perhaps there is joint ability to do something by such a group if the individual agents could put themselves into position to do it, say after sufficient discussion when there is time, or if they already are in such position because only a single pattern of cooperation is salient.[32]

It seems plausible that these conditions suffice to mandate whatever cooperative activity is required to defeat catastrophic climate change. If so, when such change occurs, it will be true of us that we have failed in consequential terms to do what we should and could have done. But that will be a problem with us, and not with act consequentialism. It will be our self-defeat, but not the self-defeat of that theory, when it is developed in the proper form.

There remains a set of conflicts for act consequentialism. Traditional individual act consequentialists may not always be able to cooperate in beneficent group acts, partly because it conflicts with their individual pursuit of the good and partly because they may not seem to others to be sufficiently trustworthy cooperators. And there may be a variety of different groups of possible cooperators available, depending partly on the terms of the cooperation. So there may be conflicts in deciding which group to join up with. These are conflicts among the merely possible group acts in which we might participate.

What are we to do about these conflicts? I believe that we are to balance the alternatives by the mechanism of multiple-act consequentialism. We should join novel group acts by regard to how much our individual contribution will add to the positive consequences of that act, but once we are inside then we should defect from such a group act only to a dominant group or individual act, an act which has more weighty positive consequences. So the ratchet mechanism of multiple-act consequentialism will shove us inside certain nascent group acts and then keep us there in many important cases. This will allow others to trust us sufficiently so that they can engage in weighty beneficent cooperation with us, and allow us to honestly join some consequentially weighty but not strictly optimific cooperative schemes, schemes that are not the very best that the group as a whole could, acting together, possibly achieve. And yet we will remain act consequentialists of a type, because of the consequentially weighty nature of the relevant beneficent group acts and because of the consequentially sensitive way we choose which group acts to join.

Still, this leaves some of the conflicts we have noted apparently unresolved. What of our act consequentialist villagers under the boulder? How does this mechanism help them, when they have so little time to generate a scheme for cooperation and no salient pattern for that cooperation?

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Unfortunately, it is at least possible that the world conspires, in the manner of the playful Martians, to force self-defeat on act consequentialism whatever modifications (p. 527) we introduce. There is no escaping that. But in reality we can generate appropriate cooperative schemes, novel group acts, for foreseeably threatening emergency situations. If act consequentialist villagers live in such a terrain, with the real possibility of boulders raining down from mountains above, they could and should predesignate a decider for such an emergency situation. They will then participate in a group act of accepting such an emergency authority that will be very beneficent should emergencies arise. What of the conflicts involving time of our act consequentialist baker, which we left hanging earlier? Individuals who will face choices like that involved in baking the cake, as we all do, should predesignate an appropriate decider as well, often the time slice of the individual at the beginning of the process, and then they will be bound inside weighty ongoing projects by the mechanism of multiple-act consequentialism.

And so there is a moderately happy ending. No plausible moral theory can avoid self-defeat of a kind in at least hypothetical circumstances. But we have seen that the apparent conflicts inherent in act consequentialism in realistic situations can be resolved, and by mechanisms that provide it with much more intuitive normative implications than it would otherwise have, especially because of the consequentially weighty effects of beneficent and cooperative group acts.[33]

References

Åqvist, Lennart. 1969. “Improved Formulations of Act-Utilitarianism.” *Nous* 3: 299–327.

Bergström, Lars. 1966. *The Alternatives and Consequences of Actions*. Stockholm: Almqvist and Wiksell.

Bratman, M. E. 1992. “Shared Cooperative Activity.” *The Philosophical Review* 101: 327–341.

Castañeda, Hector-Neri. 1968. “A Problem for Utilitarianism.” *Analysis* 28: 141–142.

Feldman, Fred. 1980. “The Principle of Moral Harmony.” *The Journal of Philosophy* 77: 166–179.

Feldman, Fred. 1986. *Doing the Best We Can*. Dordrecht, the Netherlands: Reidel.

Gibbard, Allan. 1965. “Rule-Utilitarianism: Merely an Illusory Alternative?” *Australasian Journal of Philosophy* 43: 211–220.

Gibbard, Allan. 1978. “Act-Utilitarian Agreements.” In *Values and Morals*, edited by A. Goldman and J. Kim, 91–119. Dordrecht, the Netherlands: Reidel.

Gibbard, Allan. 1990. *Utilitarianism and Coordination.* New York: Garland. (Reprint of 1971 dissertation).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Gibbard, Allan, and Harper, William. 1978. “Counterfactuals and Two Kinds of Expected Utility.” In *Foundations and Applications of Decision Theory*, edited by A. Hooker, J. J. Leach, and E. F. McLennen, 125–178. Dordrecht, the Netherlands: Reidel.

Gilbert, Margaret. 1996. *Living Together*. Lanham, MD: Rowman & Littlefield.

Gilbert, Margaret. 2006. *A Theory of Political Obligation.* Oxford: Clarendon Press.

Goldman, Holly S. 1976. “Dated Rightness and Moral Imperfection.” *The Philosophical Review* 85: 449–487.

Goldman, Holly S. 1978. “Doing the Best One Can.” In *Values and Morals*, edited by A. Goldman and J. Kim, 185–214. Dordrecht: Reidel.

(p. 528) Gruzalski, Bart. 1981. “Utilitarian Generalization, Competing Descriptions, and the Behavior of Others.” *Canadian Journal of Philosophy* 11: 487–504.

Hare, Caspar. 2013. *The Limits of Kindness.* Oxford: Oxford University Press.

Harrison, J. 1952–1953. “Utilitarianism, Universalisation, and Our Duty to Be Just.” *Proceedings of the Aristotelian Society* 53: 105–134.

Harrod, R. F. 1936. “Utilitarianism Revised.” *Mind* 45: 137–156.

Hodgson, D. 1967. *Consequences of Utilitarianism.* Oxford: Clarendon Press.

Jackson, F., and Pargetter, R. 1986. “Oughts, Options, and Actualism.” *The Philosophical Review* 95: 233–255.

Kagan, Shelly. 2011. “Do I Make a Difference?” *Philosophy & Public Affairs* 39: 105–141.

Lewis, David. 1973. *Counterfactuals*. Cambridge, MA: Harvard University Press.

Mendola, Joseph. 1986. “Parfit on Directly Collectively Self-Defeating Moral Theories.” *Philosophical Studies* 50: 153–166.

Mendola, Joseph. 1987. “The Indeterminacy of Options.” *American Philosophical Quarterly* 24: 125–136.

Mendola, Joseph. 2006. *Goodness and Justice.* Cambridge: Cambridge University Press.

Mendola, Joseph. 2014. *Human Interests.* Oxford: Oxford University Press.

Nozick, Robert. 1969. “Newcomb’s Problem and Two Principles of Choice.” In *Essays in Honor of Carl G. Hempel*, edited by Nicholas Rescher, 114–146. Dordrecht, the Netherlands: Reidel.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Pinkert, Felix. 2014. “What We Together Can (Be Required to) Do.” *Midwest Studies in Philosophy* 38: 187–201.

Pinkert, Felix. 2015. “What If I Cannot Make a Difference (and Know It).” *Ethics* 125: 971–998.

Portmore, Douglas. 2011. *Commonsense Consequentialism.* New York: Oxford University Press.

Postow, B. C. 1977. “Generalized Act Utilitarianism.” *Analysis* 37: 49–52.

Regan, Donald. 1980. *Utilitarianism and Co-operation.* Oxford: Clarendon Press.

Ross, Jacob. 2012. “Actualism, Possibilism, and Beyond.” In *Oxford Studies in Normative Ethics*, vol. 2, edited by M. Timmons, 74–97. Oxford: Oxford University Press.

Tuomela, R. 1995. *The Importance of Us*. Stanford, CA: Stanford University Press.

Williams, Bekka. 2014. “Review of *Human Interests*.” *Notre Dame Philosophical Reviews* 2014.12.08.

Notes:

([1]) Lewis (1973), 93–94; Mendola (1987); Hare (2013), 206–212; Mendola (2014), 13–39.

([2]) Bergström (1966); Castañeda (1968).

([3]) Goldman (1976).

([4]) Jackson and Pargetter (1986).

([5]) For relevant variety in views about alternatives in these cases, see Åqvist (1969); Goldman (1978); Feldman (1986); Portmore (2011); and Ross (2012).

([6]) Feldman (1980).

([7]) Gibbard (1965); Regan (1980), 18.

([8]) Does this case meet our idealization that there is full knowledge and determinacy of alternatives? I believe that on the proper conception of alternatives it does, but even if it fails to meet that constraint, it is a significant case faced by act consequentialism.

([9]) For an argument to this conclusion, see Regan (1980), 54–55, and Parfit (1984), 54.

([10]) Mendola (1986). If you think it is relevant that the Martians aren’t doing what act consequentialism demands, then let their role be played by some diabolical machine.

([11]) This is true whatever the others do.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([12]) It is a bit like meeting Death on the road to Damascus. You can't win by taking another road, since Death will go wherever you decide to go. See Gibbard and Harper (1978). It is a bit like the situation in Newcomb's Problem: You can have either just box A or instead both boxes A and B. Box B is transparent and you can see $1,000 inside. But before your choice, an omniscient or near-omniscient predictor has predicted whether you will take just A or both A and B. If they predicted that you will take just A, they will have placed a million dollars in the opaque box A. If they predicted that you will take both boxes, they will have placed nothing in that opaque box. See Nozick (1969).

([13]) Hodgson (1967).

([14]) Though this requires relaxing some of our idealizations. See Gibbard (1978).

([15]) Gibbard (1990), 6–8.

([16]) Harrod (1936); Harrison (1952–1953). There is considerable dispute about the kind of similarity between actions that is relevant. See, for instance, Gruzalski (1981).

([17]) A clause in the ideal rules allowing defections to avert disasters is not sufficient to dispose of all pressing cases of this type.

([18]) Feldman (1980), 171–174.

([19]) Postow (1977).

([20]) Such specificity can also be delivered by rule consequentialism, if there is but one set of ideal rules, but we have already noted a serious problem with this consequentialist mechanism.

([21]) For different views of the nature of group acts, see Bratman (1992); Tuomela (1995); Gilbert (1996); and Mendola (2006), 32–42.

([22]) Mendola (2006), 64–102.

([23]) Mendola (2006), 23–63, and Mendola (2014), 256–280. Criticisms of this proposal in Williams (2014) depend on intuiting the controversial claim that it is appropriate not to vote in an election when you can foresee that the best candidate will be elected without your vote and you can gain a little positive good on the side, in other words on a controversial view of the relevance of surplus cooperation, and also on a misunderstanding of the mechanism of defection from group acts.

([24]) Mendola (2014), Part III.

([25]) Kagan (2011).

([26]) Pinkert (2015).

([27]) Mendola (2014), 361–392.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([28]) Gilbert (2006).

([29]) Regan (1980).

([30]) Pinkert (2015), 982.

([31]) Regan (1980), 124–189, develops an elaborate proposal regarding the proper decision procedure for identifying cooperators. This is arguably part of cooperative consequentialism as originally proposed, and would have consequential costs analogous to those that plague modally robust act consequentialism. But I have focused on a stripped-down form of the view that reflects our idealizing assumptions.

([32]) Pinkert (2014).

([33]) Thanks to Tom Carson and Doug Portmore for comments.

Joseph Mendola

Joseph Mendola is Professor of Philosophy at the University of Nebraska—Lincoln. His research interests include ethics, metaphysics, and philosophy of mind. He is the author of four books: Human Thought (Kluwer, 1997), Goodness and Justice (Cambridge University Press, 2006), Anti-Externalism (Oxford University Press, 2008), and Human Interests (Oxford University Press, 2014).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Consequentialism, Virtue, and Character

Julia Driver

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.22

Abstract and Keywords

This essay argues that consequentialist theories can both accommodate virtue evaluation, and, indeed, the most plausible versions must do so, and that consequentialist theories can also be structured as forms of virtue ethics. Different strategies available to the consequentialist are presented and criticized, including indirect strategies which argue that the right action is the action that the virtuous person would perform. The best way for the consequentialist to approach virtue is as another form of moral evaluation understood in consequentialist terms which is distinct from consequentialist act evaluation; that is, evaluating action is only one part of an overarching consequentialist account of moral evaluation, and the theory can also be applied to dispositions, motives, intentions, etc.

Keywords: consequentialism, virtue, virtue ethics, consequentialist theories, act evaluation

WHAT can a consequentialist say about moral virtue? Quite a bit, as it happens. One impetus for the development of virtue ethics was a dissatisfaction with impartial theories such as utilitarianism. It was thought that the theory could not accommodate partial norms, such as those we have with respect to friends and family, and that it would demand that the good moral agent renounce certain virtues such as integrity (Badhwar 1991; Williams 1973).[1] Consequentialism refers to a group of approaches to moral evaluation and, in some cases, decision procedures, that includes utilitarianism but is broader. To some extent, dissatisfaction with utilitarianism in its classical formulation has led to the development of other consequentialist alternatives, all of which have been termed "indirect," though they are indirect in two very distinctive ways. Some are indirect in the sense that they hold that "right" action is a function of consequences produced by *some other thing*, such as a set of rules. Some are "indirect" in the sense that they hold that one can be a good person by consequentialist standards, even if one is not directly guided by consequentialist considerations. Part of this picture will involve an account of moral virtue, an account of good dispositions. Most, including myself, have opted for the second strategy, though later in this essay we will discuss a virtue consequentialism that is an example of the first strategy. First, however, we need to make important distinctions.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(p. 464) 1. Virtue Theory and Virtue Ethics

We need to distinguish two claims:

1. Consequentialism has no place for virtue evaluation.
2. There is no such thing as a consequentialist virtue ethics.

Both of these claims are false. The first does not necessarily reference virtue *ethics*. Virtue ethics holds that virtue evaluation is in some way "primary" to moral evaluation. For example, a virtue ethics that makes room for deontic act evaluations would hold that the right action is understood in terms of virtue, not that virtue is understood in terms of right action. The right action is what the virtuous person would do, or is what virtue demands in the circumstances, and so on. But one can be a critic of virtue ethics and still maintain that virtue evaluation is important, that it plays an important role in our critical practices, and that it might not be understood in terms of right action. In *Utilitarianism*, John Stuart Mill argued that the good utilitarian agent would desire virtue "for itself" and that virtue evaluation had an important place in a comprehensive utilitarian theory. In his *A System of Logic*, in the section "Art of Life," he discusses the importance of cultivating virtue (Driver 2015). Roger Crisp argues that utilitarianism, understood properly, recommends a life of virtue with the correct principle of evaluation operating as a kind of regulative ideal (Crisp 1992).

As I will discuss in more detail, the second claim is also false. One could understand right action in terms of virtue, for example, while maintaining a consequentialist account of virtue. Accounts of virtue are part of virtue *theory*, and they may or may not be tied to a particular virtue ethics. Both of these options will be discussed herein.

Regarding the first claim, the criticisms from neo-Aristotelians and anti-theorists went beyond simply noting that contemporary consequentialists didn't seem to make a place for virtue. Even if one could provide a consequentialism with virtue evaluation as a component, it would be an inadequate account. This criticism rested on the view that the characteristic practical deliberations of a virtuous person would explicitly make reference to the justifying reasons for the action in question. Otherwise, as Michael Stocker pointed out, the virtuous agent would be "alienated" from the norms governing his own action. This would be very strange—and other writers found this odd, as well. Indeed, later writers such as Alison Hills explicitly held that the virtuous agent must not only act rightly, but do so out of moral understanding, which requires a systematic understanding of moral norms, and certainly an understanding of those that justified the virtuous agent's actions. Thus, the virtuous consequentialist would presumably be guided by maximizing the good, impartially considered. And this is clearly incompatible with many types of virtue, such as the virtues of being a good parent, partner, or friend. When someone does something good for a friend, it is not to maximize the good—indeed, it is likely to conflict with maximizing the good in that instance.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(p. 465) However, consequentialists such as Peter Railton challenged this criticism by arguing that sophisticated consequentialism, one which holds that being disposed well isn't necessarily or likely to involve always thinking about promoting the good, is fully able to provide a plausible account of these good dispositions, or virtues (Railton 1984). On his view, it makes good sense on consequentialist grounds for someone to cultivate virtues of partiality, such as dispositions to favor the near and dear. Love and friendship are relationships that are part of a good, happy, human life. These relationships have enormous value.[2] It is thus very much in keeping with consequentialism to grant them an important role in the promotion of human happiness.

Railton was presenting a form of objective consequentialism in which the moral quality of actions and character traits is determined by whether or not they were actually good producing. Given some very plausible empirical assumptions, trying to promote the good in every instance is unlikely to actually promote the good. Instead, we need good-producing dispositions, which themselves may not be characterized by the sort of practical deliberation critics of consequentialism found so troubling. In *Uneasy Virtue* (Driver 2001), my aim was to develop such an account further by arguing that moral virtues are systematically productive of the good, and that this did not require that the virtuous agent be thinking in terms of producing the good.

Part of the aim was to show that consequentialist accounts of evaluation are fully able to incorporate virtue evaluation. Indeed, my view is that such accounts do a better job than the very popular neo-Aristotelian accounts of virtue. Neo-Aristotelian accounts have, famously, placed high demands on virtue, such that they seem to generate implausible verdicts about virtue. For example, Aristotle held that the virtuous person knows what he is doing under the description that renders the act an expression of virtue. This seems quite true for very many virtues. For example, generosity requires that the generous person realize that she is helping someone in need. However, it doesn't seem true for all virtue. One example is the virtue of modesty, of underestimating one's worth to some small or moderate degree. Since this involves epistemic error, it would be disqualified on Aristotle's account. Further, some virtuous individuals possess virtue even when they have a mistaken view of what morality demands. In *Uneasy Virtue*, I discussed the case, made famous by Jonathan Bennett, of Huckleberry Finn who, out of sympathy for his friend Jim, an escaped slave, does not turn Jim in to the authorities even though he believes that helping Jim is morally wrong (Bennett 1974). If the virtuous person must be acting in light of a conception of the good, as many Aristotelians believe, then Huckleberry's sympathy is not virtuous.

One way to avoid these sorts of difficulties would be to develop a consequentialist account of virtue that did not require that virtuous agents view themselves as doing good. In the case of modesty, we might hold that the modest person just doesn't rank. It isn't that the modest person thinks about ranking and then just decides not to rank. It doesn't even occur to him. Thus, the standard for virtue should be psychology-insensitive. (p. 466) This suggests that the best approach is a version of *objective consequentialism* earlier discussed by Railton. When it comes to act evaluation, the objective consequentialist

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

holds that the right action is the one that produces the best overall consequences. It is contrasted with subjective consequentialism, which can be unpacked in two ways: as psychology-sensitive or as evidence-sensitive. These may overlap, but they need not. The psychology-sensitive versions take a variety of forms. For example, the right action is the action that the agent believes will produce the best consequences, where the agent's empirical beliefs and beliefs about the good are taken as authoritative. This standard is not very plausible. But partially psychology-sensitive standards are much more plausible. For example, Frank Jackson has the view that the right action is a function of what the agent *actually* believes and what the agent *ought* to desire (Jackson 1991).

However, the standard can also be evidence-sensitive. For example, we might hold that the right action is the action that promotes the good relative to the evidence available to the agent. If the agent is not availing herself of the evidence, then this would not be psychology-sensitive. In *Uneasy Virtue*, I was more concerned with developing an account that was not psychology-sensitive, since I wanted to account for virtues like modesty and Huck's brand of sympathetic engagement. Thus, I opted for the objective account: virtues are character traits that *systematically* produce more good than not. The "systematically" is required to insulate this form of evaluation from the problem of moral luck. Further, on the view that I proposed in a later book, *Consequentialism*, the subjective standards offered standards of praise and blame, rather than a standard of right. So, for example, the right action is the one that promotes the good, though one only warrants praise if one performs the action because one expects it to promote the most good given one's evidence (where that is not understood as an explicit expectation, but simply recoverable from the agent's psychology in some way).

This consequentialist account of virtue has been roundly attacked in the virtue literature. Some of the criticisms focused on the account of modesty, but many also centered on the overall view of moral virtue. Primarily, the criticisms focused on the problems relating to abandoning a psychology-sensitive standard. One line of attack is to hold that since virtues are human excellences they cannot be *modally fragile*. Since the objective consequentialist view is externalist in nature, it is committed to the view that if one changes the external circumstances sufficiently, then a trait that was a virtue in some other context would no longer be a virtue. This particular version of the modal fragility problem would be common to any account that held that actual consequences produced by a trait are relevant to that trait's status as a virtue. The more serious problem has to do with the view implying that traits which seem intrinsically awful to us since they are characterized by bad intentions or desires (e.g., a desire to harm) might conceivably be part of a virtue simply because the consequences would be good enough. Of course, this imagines a world very different from our own, but it is not conceptually or physically impossible. This does seem to be a problem for a straightforward objective account, because if we are going to allow for intrinsically bad desires, then we are at least violating the spirit of an externalist approach. However, later in this essay I will offer a modification (p. 467) of my original account that allows the objectivist to exclude problematic intentions and desires.[3]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Ben Bradley has argued that the account cannot adequately handle moral luck cases, another way to look at modal fragility. He argues for the following:

> VC: It is a virtue for people S1-Sn to have character trait V1 rather than character trait V2 at world *w* iff (i) V2 is a member of the contrast class of V1, and (ii) the expected intrinsic value of a closest world to *w* where S1-Sn exercise V1 is greater than the expected intrinsic value of a closest world to *w* where S1-Sn exercise V2. (Bradley 2005, 295)

An example Bradley uses to illuminate the account is that of Lucky. Lucky is a psychopath, and an active one. He is always trying to kill someone. However, due to sheer bad luck, whenever he tries to kill someone, the attempt fails due to some external intervention (wind nudges the bullet away, the poison is diluted by a rainstorm, someone readjusts the torn carpet at the top of the stairs, etc.). In spite of the fact that his murderous intentions never lead to anything bad, we would certainly want to say that Lucky is vicious, that he is, for example, malicious. Bradley notes that as stated this isn't a counterexample to my view since I want to allow that these attributions are tied to systematic production of the good, and Lucky is just a one-off case. However, he notes that we could concoct scenarios in which all the malicious people in the world were systematically unlucky when they attempted to exercise their malice. What the contrastive account does is provide a better way to insulate agents from moral luck. In the case of Lucky, the expected intrinsic value of a world in which is malicious is much lower than the expected intrinsic value of a world in which he is not. This is because the unlucky Lucky will succeed in close-by possible worlds, and this affects what we are entitled to expect. While I don't exactly agree with Bradley's account, I do believe that some versions of modal fragility pose genuine problems that need to be addressed.

Another problem that has been levied against this account holds that it is just too thin (Calder 2007). It doesn't provide enough detail for identifying virtues. However, in *Uneasy Virtue*, I argued that when it came to particular virtues, all sorts of different factors should be taken into consideration. Just because good intentions are not necessary for a virtue like modesty does not mean that they are not necessary for a virtue like generosity. This is because in order for the trait to be generosity it must involve acting with the intention of helping those in need, which is a paradigmatically good intention. My claim was simply that there are virtue traits that do not require this (though perhaps they do require the absence of a bad intention).

(p. 468) Robert Adams criticizes the view by granting that the virtues are conducive to good consequences, but he holds that this doesn't capture the full nature of virtue (Adams 2006). Instead, what is also required is that virtuous people are those who are *for the good*. Being for the good is an essential component, and it is intrinsically important, not simply important for instrumental reasons, which a consequentialist could easily grant. However, Adams distinguishes between motivational virtues, like generosity, and structural virtues such as courage and conscientiousness. As noted earlier, it seems very plausible that generosity does require being for the good in the form of helping those in

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

need. But this isn't clear with the structural virtues, and Adams himself claims that someone can be genuinely courageous, that is, have the virtue of courage, without having good aims (Adams 2006). Other writers, such as Elinor Mason, argue that traits like conscientiousness don't *operate as* virtues unless the agent's aims are good ones (Mason 2019, 80–81). Structural virtues, do, however, concern *certain* goods—a well-organized life, a life of commitment to certain values, and so on. A fully virtuous person will have structural virtues. On my view Adams's account will be insufficient to capture the wide range of virtues that have historically been identified as virtues, or human excellences. One example that I will discuss next concerns *artificial virtues*, such as justice.

A possible solution to some of these problems is to hold that there is *some* responsiveness to the right reasons that underlies virtue, but that it is much thinner than that required by neo-Aristotelians.[4] Indeed, for most virtues this seems entirely true. Nomy Arpaly argues for a view such as this, though with more robust requirements, in *Unprincipled Virtue* (Arpaly 2002a). However, I think that a better way to put it is to hold that the virtuous person at least is not improperly responsive to the wrong reasons. This, again, requires me to reject something that I argued for in *Uneasy Virtue*. While it may be true that we can value people who can do difficult things, that require a certain kind of meanness, when it is channeled properly and leads to good (e.g., imagine a society in which ruthlessness that didn't actually result in overall harm was necessary on the part of some in order to bring about a very large benefit), we cannot call them virtuous if their motives are actually *bad* motives. They may not need to have good motives, true, but they can't actually be bad; they cannot actually aim for the bad.[5]

This leads to the possibility that some virtues can be characterized as involving *absences* of bad-making features (Driver unpublished manuscript). The best way to illustrate this is by using cases. The best cases involve people who are alienated from the value system they were raised in. Imagine someone two hundred years ago who believed that killing animals for food was morally wrong. Unlike their peers, they don't see the animals' lack of rationality as being relevant to this issue. Thus, they are not accepting as a reason the wrong reason that other people accept. We can assume that their desires to (p. 469) act rightly and avoid wrongdoing are the same as everyone else's, de dicto. Otherwise the path to virtue would be too easy, involving a simple intellectual exercise.

One could, of course, reject an objective consequentialist account and instead opt for some type of subjective version, either a psychology-sensitive or evidence-sensitive account. For example, one might hold that a virtue is a tendency to do the best one can, given the evidence one has. Specific virtues would then be delineated in terms of what they are *about*. Of course, the objective consequentialist can take this on-board as a standard of praise and blame. Thus, one could still be an objective consequentialist but argue that virtue is tied to the standard of praise and blame.

This is the view that I now favor, as long as we include in the standard of praise and blame a very *thin* notion of responsiveness to the right reasons and a failure to be moved by the wrong reasons. This would handle most of the cases that I was worried about in

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

my earlier work. Even so, this still presents us with a possible problem. Some virtues don't even seem to require *this* (Driver 2016). Consider Hugo Grotius on injustice:

> …injustice is nothing else in its nature than the usurpation of what is another's; nor, does it make any difference whether that proceeds from avarice, or from lust, or from anger, or from thoughtless compassion. (Grotius [1625] 1853)

To be just, then, is to simply respect the law. One might argue that Grotius is using the word "justice" as a deontic term rather than as an aretaic term. And there is no doubt that Grotius seems to go further than other theorists in minimizing what is required for virtue. To be fair, Grotius isn't offering a fully worked-out theory. If he did, he might have his own version of the modal fragility problem. But we needn't go as far as Grotius to note that there seems to be an important difference between certain virtues such as benevolence, and others such as justice, and that distinction can relate to the acceptable motives of each. Indeed, this is a long-standing distinction in the history of ethics. This at least suggests that a person can possess the virtue of justice without being motivated by a concern for the well-being of others. Perhaps there is simply an attitude in favor of justice itself, or something associated with it such as equality, fairness, and so on.

Another account of virtue that is *compatible with* consequentialism is one proposed by Thomas Hurka. On Hurka's view virtue is characterized by the appropriate attitudes toward good and evil. The virtuous person loves the good and hates evil. The reason this account is compatible with consequentialism is that virtue is still defined in terms of the good; even though it is not *simply* instrumentally valuable, it is intrinsically valuable. Like Adams's account, virtue is understood in terms of having some pro-attitude toward the good. However, Hurka also adds the attitude must be proportional. If Abigail really, really, despises rude driving, to the point of becoming extremely angry, then her attitude of dislike toward the rudeness might be out of proportion and not virtuous. Adams believes that this renders Hurka's account implausible, since it would seem to require that we love everyone equally since all lives have the same objective value. As Adams notes, however, Hurka would resist this characterization as he does allow for agent-relative (p. 470) value in the sense that he thinks that we can having differing reasons that will lead us to appropriately favor one good over another.

2. How Would One Develop a Consequentialist *Virtue Ethics*?

At the opening of this essay, I also claimed that one could develop a consequentialist virtue ethics.[6] One possibility would be to develop it as a form of ethics that eliminates deontic act evaluations altogether. We don't need an account of right action in terms of virtue because we don't need an account of right action at all. There might be an Anscombian justification for this (Anscombe 1958). Virtue and vice terms are positively and negatively valenced, but as thick concepts they also include more information. Do I need to know anything other than that someone is doing something *unjust*? The thinner, deontic

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

terms are simply vestiges of a system of morality that is no longer dominant. But this strikes many as too extreme, as well as just false. It is important to be able to distinguish right actions from virtuous ones, and wrong actions from vicious ones. For one thing, many find the distinction between "right action" and "morally worthy action" to be illuminating and important (Arpaly 2002b; Hills 2009; Markovits 2010). A more promising approach is to keep morally "right" and "wrong" but in a way that gives virtue explanatory primacy within the theory.

Rosalind Hursthouse develops a neo-Aristotelian virtue ethics that defines right action in terms of the virtues, where the virtues are unpacked along Aristotelian lines (Hursthouse 1999). She writes that:

> An action is right iff it is what a virtuous agent would, characteristically, do in the circumstances. (Hursthouse 1999, 79)

She includes exceptions for tragic dilemma cases, in which even the best course of action doesn't seem right, and notes that there are two senses of "right" that we often use: the sense in which the right action is "the thing to do" and the sense in which the right action warrants a "tick of approval." In tragic dilemmas there may be a right thing to do in the (p. 471) first sense but not in the second. But the basic idea is that the right is understood in terms of the virtuous. Thus, the virtuous has "priority."

One could adopt a similar *structure* and instead adopt a consequentialist account of the virtues. This form of indirection would be similar to rule consequentialism and would have many of the same problems. Either the action the virtuous person would perform could be suboptimal, but still "right," which seems irrational, or it is "right" simply in virtue of it being what the virtuous person would do, which is arbitrary. Further, as in Hursthouse's own account, it isn't clear that it is plausible to think that what any *particular* agent in a certain set of circumstances should do is what the virtuous person in those circumstances would do. If Lindsay, who is a bit of a coward when it comes to water, were to try to rescue someone drowning, that might be a disaster even though it is what the *virtuous* person would do.

Another approach to virtue ethics, developed by Michael Slote, is what he terms an *agent-basing approach* in which right action is understood as an expression of morally good motivational structure (Slote 2001). On this view, when a person acts in such a way as to express her virtuous motivational system, the action is right. One could develop a consequentialist version of this approach, one which holds that right action is an expression of an agent's good motivational structure, where that is understood in consequentialist terms. Indeed, this would be similar to what Robert Adams has termed "motive utilitarianism" (Adams 1976). Most of the work would be done by specifying what constitutes a morally good motivational structure on consequentialist grounds. If one wanted to pursue this strategy, then it could be integrated with approaches such as Peter Railton's account from "Alienation, Consequentialism, and the Demands of Morality." Railton himself distinguishes right action from good dispositions, but one could imagine modifying the view that having a good dispositional structure might sometimes mean deviating from

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

the right, to the view that the right just is the *expression of* that structure. A well-motivated action—that is, the action done for the right reasons where "right reasons" simply refers to the quality of the agent's will or motivation—just is the right action. This would have the same set of problems that any version of agent-basing has. A noncircular account of "expressing" would need to be developed. If it simply means that the good motivations cause the right action, then the account runs the risk of mischaracterizing some actions as right simply in virtue of the fact that they were caused in some odd way by the good motivational set. If "expression" is more than mere "cause," then it incorporates reference to right-making features of the action, and the right no longer depends upon simple expression of good motivation (Driver 1995). Also, like Hursthouse's account, it would seem to allow for a wide variety of "right" actions, which seems counterintuitive when we think that at least in some cases there clearly is *the* right thing to do even among well-motivated options.[7]

References

Adams, Robert. 1976. "Motive Utilitarianism." *The Journal of Philosophy* 73, no. 14: 467–481.

Adams, Robert. 2006. *A Theory of Virtue: Excellence in Being for the Good*. New York: Oxford University Press.

Annas, Julia. 2011. *Intelligent Virtue*. New York: Oxford University Press.

Anscombe, G. E. M. 1958. "Modern Moral Philosophy." *Philosophy* 33, no. 124: 1–19.

Arpaly, Nomy. 2002a. *Unprincipled Virtue*. New York: Oxford University Press.

Arpaly, Nomy. 2002b. "Moral Worth." *Journal of Philosophy* 99, no. 5: 223–245.

Arpaly, Nomy, and Schroeder, Timothy. 2013. *In Praise of Desire*. New York: Oxford University Press.

Ashford, Elizabeth. 2000. "Utilitarianism, Integrity, and Partiality." *Journal of Philosophy* 97, no. 8: 421–439.

Badhwar, Neera K. 1991. "Why it is Always Wrong to be Guided by the Best: Consequentialism and Friendship." *Ethics* 101, no. 3: 483–504.

Bennett, Jonathan. 1974. "The Conscience of Huckleberry Finn." *Philosophy* 49, no. 188: 123–134.

Bradley, Ben. 2005. "Virtue Consequentialism." *Utilitas* 17, no. 3: 282–298.

Bradley, Ben. 2016. "Character and Consequences." In *Questions of Character*, edited by Iskra Fileva, 78–88. New York: Oxford University Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Calder, Todd. 2007. “Against Consequentialist Theories of Virtue and Vice.” *Utilitas* 19, no. 2: 201–209.

Cohon, Rachel. 2008. *Hume’s Morality: Feeling and Fabrication*. New York: Oxford University Press.

Crisp, Roger. 1992. “Utilitarianism and the Life of Virtue.” *Philosophical Quarterly* 42, no. 167: 139–160.

Driver, Julia. “Absence and Virtue.” Unpublished manuscript.

Driver, Julia. 1995. “Monkeying with Motives: Agent-Basing Virtue Ethics.” *Utilitas* 77, no. 2: 281–288.

Driver, Julia. 2001. *Uneasy Virtue*. New York: Cambridge University Press.

Driver, Julia. 2015. “Mill, Sentimentalism, and the Cultivation of Virtue.” In *Cultivating Virtue: Perspectives from Philosophy, Theology, and Psychology*, edited by Nancy E. Snow, 49–64. Oxford: Oxford University Press.

Driver, Julia. 2016. “Minimal Virtue.” *The Monist* 99, no. 2: 97–111.

Garrett, Don. 2007. “The First Motive to Justice: Hume’s Circle Squared.” *Hume Studies* 33, no. 2: 257–288.

Grotius, Hugo. 1853. *On the Rights of War and Peace*. Trans. William Whewell. Cambridge: Cambridge University Press.

Hills, Alison. 2009. “Moral Testimony and Moral Epistemology.” *Ethics* 120, no. 1: 94–127.

Hurka, Thomas. 2001. *Virtue, Vice, and Value*. New York: Oxford University Press.

Hursthouse, Rosalind. 1999. *On Virtue Ethics*. New York: Oxford University Press.

Jackson, Frank. 1991. “Decision-theoretic Consequentialism and the Nearest and Dearest Objection.” *Ethics* 101, no. 3: 461–482.

Lord, Errol. 2017. “On the Intellectual Conditions for Responsibility: Acting for the Right Reasons, Conceptualization, and Credit.” *Philosophy and Phenomenological Research* 95, no. 2: 436–454.

Markovits, Julia. 2010. “Acting for the Right Reasons.” *Philosophical Review* 119, no. 2: 201–242.

(p. 473) Mason, Elinor. 2019. *Ways to Be Blameworthy*. New York: Oxford University Press.

Railton, Peter. 1984. “Alienation, Consequentialism, and the Demands of Morality.” *Philosophy and Public Affairs* 13, no. 2: 134–171.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Russell, Paul. 2013. “Hume’s Anatomy of Virtue.” In *The Cambridge Companion to Virtue Ethics*, edited by Daniel C. Russell, 92–123. New York: Cambridge University Press.

Skorupski, John. 2001. “Externalism and Self-Governance.” *Utilitas* 16, no. 1: 12–21.

Slote, Michael. 2001. *Morals from Motives*. New York: Oxford University Press.

Swanton, Christine. 2003. *Virtue Ethics: A Pluralistic View*. New York: Oxford University Press.

Taylor, Jacqueline. 2002. “Hume on the Standard of Virtue.” *The Journal of Ethics* 6, no. 1: 43–62.

Van Norden, Bryan. 2007. *Virtue Ethics and Consequentialism in Early Chinese Philosophy*. New York: Cambridge University Press.

Williams, Bernard. 1973. “A Critique of Utilitarianism.” In *Utilitarianism: For and Against*, J. J. C. Smart and Bernard Williams, 77–150. New York: Cambridge University Press.

Notes:

([1]) For an argument undercutting Williams on this issue, see Ashford (2000).

([2]) Further, it is open to the consequentialist to view these relationships as having intrinsic value, which would eliminate most of the instrumentalist worries.

([3]) John Skorupski discusses a non-modal version of a similar problem: our access to evidence is highly restricted to our own little corner of the universe (Skorupski 2001). Suppose that whenever Sandy tried to help someone an evil demon hurt two other people somewhere else, and that the demon does this whenever anyone tries to help. A consequentialist, it seems, would be committed to holding that trying to help those in need is not a virtue. The strategy for dealing with this problem will be the same as the strategy with modal fragility.

([4]) Errol Lord has provided an account that doesn’t fall prey to the typical criticisms of the Aristotelian approach (Lord 2017).

([5]) There are other ways one might want to spell this out. For example, Nomy Arpaly and Timothy Schroeder view virtues as boiling down to intrinsically good *desires* (Arpaly and Schroeder 2013).

([6]) There are many different approaches to virtue ethics today. See Annas (2011) and Swanton (2003). For some reason, a full Humean virtue ethics has not emerged, though there are very many works on Hume’s virtue *theory*; see, for example, Cohon (2008), Garrett (2007), Taylor (2002), and Russell (2013). Further, some virtue ethicists such as Slote have been influenced by features of Hume’s account, particularly the moral sentimentalism. However, given this essay is on consequentialism and virtue, I am only discussing approaches that I believe could mirror similar consequentialist approaches to virtue ethics.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Another thing to note is that debates between virtue ethicists and consequentialism appear in non-Western systems of ethics, which Bryan Van Norden explores in traditional Chinese ethical theories (Van Norden 2007).

([7]) Some of the material in this essay is drawn from my *Uneasy Virtue* (2001) and my article “Minimal Virtue” (2016).

Julia Driver

Julia Driver is Professor of Philosophy at the University of Texas at Austin and Professorial Fellow at the Centre for Ethics, Philosophy, and Public Affairs at St. Andrews. Her research is primarily focused on normative ethics, metaethics, and moral psychology. She is the author of several books, the most recent being Consequentialism (Routledge, 2012).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ConsequentIalizing

Paul Hurley

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.23

Abstract and Keywords

The strategy of consequentializing features that are intuitively relevant to the deontic evaluation of actions by building them into the telic evaluation of outcomes is almost as old as consequentialism itself. But the recent rejection by many consequentialists of the traditional commitment to an agent-neutral constraint on the relevant evaluation of outcomes has ushered in new consequentializing arguments for consequentialism and new consequentialist arguments for consequentializing. While the former fail, the latter ground the case for consequentializing in deeply entrenched and widely held commitments. These commitments to outcome-centered accounts of reasons, actions, and attitudes dictate that any plausible alternative account of what agents rationally and morally ought to do must be a form of consequentialism and hence must have a consequentialized form. Such outcome-centered commitments, however, all run afoul of common sense in similar ways, and a pervasive strategy for mitigating this counter-intuitiveness trades upon a conflation of two distinct senses in which we speak of actions as bringing about outcomes.

Keywords: consequentialism, consequentializing, agent-neutral, agent-relative, deontologizing, explanatory commitment, deontic constraints

> The main strategy for “consequentializing” any given moral theory is simple.
>
> We merely take the features of an action that the theory considers to be relevant, and build them into the consequences.
>
> —Dreier (1993), 23

1. Introduction

WHOLESALE consequentializing of the sort described in the epigraph is a recent development in the evolution of consequentialist moral theory. But the general strategy of identifying features nonconsequentialists take to be directly relevant to the deontic evaluation of actions as right and wrong, prohibited and permitted, and building them into the telic

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

evaluation of outcomes as better, worse, and best, is almost as old as consequentialism itself. What's new is the challenge by recent consequentializers to the traditional parameters within which this strategy has been deployed. Until recently, consequentialists have insisted that the relevant ranking of such outcomes must be from an impersonal/God's eye point of view of the universe—an agent-neutral point of view. Moreover, they have taken it to be obvious that on any recognizably consequentialist theory actions are morally right *because* they bring about the best outcomes. It is always right to do what brings about the best *overall* outcome, and it is right *because* it brings about the best overall outcome.

Recently, however, a new wave of consequentialists has challenged the legitimacy of the agent-neutral commitment, and many have gone further, challenging the explanatory commitment as well.[1] Rejecting either or both of these commitments opens the (p. 26) door for consequentializing many more of what are traditionally considered the nonconsequentialist features relevant to the evaluation of actions, and doing so in a way that tracks much more closely such common-sense deontological evaluations.

It will first be helpful to introduce some terminology. Let us say that a theory is in *consequentialized form* if it holds that an act merits the relevant deontic evaluation (is right or morally required or what the agent ought or morally ought to do) if and only if it brings about the best outcome.[2] Let us say that a theory is a *form of consequentialism* if, in addition, it purports to explain the relevant deontic evaluations of actions through appeal to the prior evaluation of outcomes. Consequentialized form requires only the equivalence between deontic evaluation of actions and telic evaluation of outcomes; a form of consequentialism purports to provide a plausible account of why actions have the deontic status they do through appeal to the value of outcomes.[3]

In what follows I will first briefly take up the history of the deployment of the consequentializing strategy within the traditional agent-neutral and explanatory commitments. I will then look at proposals to articulate counterparts to nonconsequentialist theories in consequentialized form without any commitment to the explanatory priority of the evaluation of outcomes. The claim that we can thus consequentialize all moral theories is often taken to establish that we ought to take every theory to be a form of consequentialism. I will refer to this in what follows as the *consequentializing argument for consequentialism*. Even if such nonexplanatory consequentializers are right that every moral theory has a counterpart in consequentialized form, I will show that it in no way follows that every theory is plausibly viewed as a form of consequentialism.

Next, I will take up arguments that abandon the agent-neutral commitment while maintaining the explanatory commitment in some form. I show that advocates of this approach have generated an argument grounded in widely held premises for the claim that any plausible theory must be a form of consequentialism and hence not only can be but must be rendered in consequentialized form. I will refer to this argument in what follows as the *consequentialist argument for consequentializing*. This argument for consequentializing is not susceptible to many of the arguments against traditional agent-neutral consequen-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

tialism, but I will sketch certain very different kinds of criticisms to which I suspect that it is susceptible. My focus throughout will be entirely upon act consequentialism, both because this is the focus of virtually all of this recent work, and because this work provides additional grounds for such a focus.

(p. 27) 2. Traditional Agent-Neutral Consequentializing

Bentham (1907) maintains that only quantity of pleasure, not quality, is directly relevant to the ranking of outcomes as better or worse overall; hence, the right action is the action that brings about the greatest quantity of pleasure. This strikes many as deeply counterintuitive. In particular, the quality of pleasures is also a feature that common sense joins nonconsequentialists in recognizing as relevant to the evaluation of actions. Mill's solution (2001, chap. 2) is to take this feature that common sense holds to be relevant to the evaluation of actions, the quality of pleasure, and build it in as a consideration directly relevant to the ranking of outcomes as better and worse overall. Pace Bentham, quality as well as quantity of pleasure is relevant in the ranking of outcomes as better or worse overall. We can understand Mill's strategy as one of consequentializing quality of pleasure.

Happiness understood as equivalent to pleasure is for Mill the only consideration that is directly relevant to the ranking of outcomes as better and worse. However, a more robust happiness that comprehends pleasure, but is not reducible to it, seems to be a feature relevant to the deontic evaluation of action. Why not then take this feature that common sense and other moral theories take to be directly relevant to the evaluation of action and build it into the ranking of outcomes? Robust happiness has been consequentialized. And so it goes. Common sense suggests that more than one feature is relevant to the evaluation of actions. Why not consequentialize a plurality of such features? In place of monism, the view that only one feature (e.g., pleasure, happiness, or well-being) is relevant to the ranking of outcomes, we now have a view that a plurality of considerations is relevant to the ranking of outcomes.[4] Common sense and alternative moral theories also suggest that intrinsically moral considerations such as rights, or Kantian respect for persons as ends in themselves, are relevant to the deontic moral evaluation of actions. If so, why not just build them directly into the consequences, ranking relevant outcomes based upon whether they minimize the extent to which rights are violated or maximize the extent to which people are treated with respect, as ends in themselves? Rights and Kantian respect for persons will have been consequentialized.[5] The history of consequentialism, it can seem, is in large part a history of consequentializing, of building into the agent-neutral ranking of outcomes more and more of the features that common sense and other nonconsequentialist moral theories suggest are relevant to the deontic evaluation of actions.

(p. 28) Until recently such consequentializing initiatives have taken place entirely within the context of the two commitments mentioned at the outset, the Agent-Neutral Commitment and the Explanatory Commitment:

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Agent-Neutral Commitment: the standpoint for the evaluation of outcomes that is relevant to the moral evaluation of actions is an impersonal, agent-neutral standpoint.

Explanatory Commitment: actions have their deontic status *because* they bring about the outcome ranked best from the relevant standpoint.

By the end of the last century, consequentialists began to chafe at the limits placed upon the consequentializing strategy by these two commitments, limits that prevented them from accommodating other central features of common-sense morality that are easily incorporated into rival moral theories. For example, common-sense morality, as well as most alternative moral theories, recognizes a fundamental role for deontic constraints, cases in which it is wrong to bring about the best outcome from an agent-neutral point of view. It is worse from an agent-neutral point of view for three promise breakings to happen than it is for me to break one promise that I have made, but common sense suggests that each of us nonetheless has distinctive moral reasons to keep one's own promises, hence that we each should keep one's promise to another even if, as a result, three others will break their promises. Three other persons will have committed those promise breakings, not me, and they should be held accountable for doing so, not me. To think otherwise would be to treat me as though I am just as responsible for their promises as I am for my own, but this flies in the face of common sense and our understanding of what we owe to each other. Common-sense morality thus includes agent-relative constraints upon bringing about the best outcome from an agent-neutral point of view, and nonconsequentialist theories readily accommodate such constraints without problematic recourse to indirection. Such agent-relative constraints prove an embarrassment to traditional agent-neutral consequentialism.

New wave consequentialists reject the Agent-Neutral Commitment, allowing them to build the *agent-relative* features of common-sense morality that generate such constraints into the ranking of outcomes. The resulting ranking of outcomes will itself be *relative to the agent*, for example, will reflect not only the value of minimizing the number of promises that are broken overall, but of minimizing the number of promises that the agent herself breaks.[6] The result of incorporating such agent-relative considerations will be an agent-relative ranking of outcomes. On such a view, it might often be the case that the best outcome relative to the agent minimizes the number of promises that she breaks, even though acting to promote this outcome will result in more promises being (p. 29) broken—a worse outcome from an agent-neutral point of view.[7] Thus, this agent-relative consequentialist account generates deontic constraints: the agent morally ought to keep her promise (promoting the best outcome relative to her) even though more promises will be broken overall (this outcome will be agent-neutrally worse).[8] Within the context of the explanatory commitment, the test of the effectiveness of this strategy will be what it has traditionally been: does a form of consequentialism that consequentializes these agent-relative features of ordinary morality provide the most plausible account of the deontic evaluation, including the deontic moral evaluation, of actions? Does it better account for what we ought and morally ought to do, and if so why? I will later show that with the re-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

jection of the agent-neutral commitment a deep argument comes into view—the consequentialist argument for consequentializing—for the conclusion that any plausible moral theory must have a consequentialized form.

Some among this new wave of agent-relative consequentializers, however, treat not just the agent-neutral commitment, but the explanatory commitment as well, as at least initially dispensable.[9] Whereas those who jettison the agent-neutral commitment but maintain the explanatory commitment are best understood as providing an argument *for* consequentializing, for insisting that all plausible alternatives be evaluated as forms of consequentialism, hence in their consequentialized forms, those who jettison both commitments sometimes seem to take the fact that every theory can be put in consequentialized form to establish that (1) every theory should be considered in its consequentialized form, and that (2) every theory should be viewed as a form of consequentialism rather than as a form of an alternative moral theory—the consequentializing argument for consequentialism.

The consequentialist argument for consequentializing and the consequentializing argument for consequentialism are not always clearly distinguished, either by their critics or by their advocates.[10] These arguments must also be distinguished from another set of arguments that I will here leave to one side, pragmatic arguments for consequentializing that are not arguments for consequentialism. Such pragmatic arguments advocate putting all theories in consequentialized form, but only for certain carefully delimited, instrumental, pragmatic purposes, purposes that do not commit advocates (p. 30) either to accepting or rejecting consequentialism itself.[11] In the next section I will take up the consequentializing argument for consequentialism before turning, in the sections to follow, to the deeper and more interesting consequentialist arguments for consequentializing.

3. The Consequentializing Argument for Consequentialism

If we bracket any need for a plausible explanatory rationale for the deontic evaluation of actions, seeking only the “if” or “if and only if,” and not the “because,” there is a straightforward formula for putting any seemingly nonconsequentialist moral theory in consequentialized form:

> Take whatever considerations that the non-consequentialist theory holds to be relevant to determining the deontic statuses of actions and insist that those considerations are relevant to determining the proper ranking of outcomes. (Portmore 2007, 39)

When such a strategy is applied to all relevant considerations, the results are taken to support a “deontic equivalence thesis”:

> For any remotely plausible nonconsequentialist theory, there is a consequentialist counterpart theory that is deontically equivalent to it[12]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

This counterpart to any nonconsequentialist theory is its consequentialized form. The claim that every plausible alternative moral theory can be put in such a consequentialized form has been challenged,[13] but even if we accept it, what follows? Some consequentializers seem to take the demonstration that theories can be put into consequentialized form to establish that it is this form in which alternative theories are best compared and contrasted with each other, and that every plausible moral theory is really a form of consequentialism, such that those who have taken themselves to be offering alternatives to consequentialism are simply mistaken.[14]

But such additional conclusions simply do not follow from the fact that such theories can be consequentialized, and arguments that they do often implicitly rely upon (p. 31) the explanatory commitment in some form. To see why, consider virtue ethics in its consequentialized form. Following Aristotle (1999), let's take virtuous action to be activity of the soul in accordance with right reason. As a person of good character, I exercise right reason through deliberation to determine what virtuous actions are. Now I put this nonconsequentialist theory in consequentialized form. I determine the most vicious act available to me and rank the outcome that results from performing such an act below all others; I then determine the most virtuous act, and rank the outcome that results from its performance above all others, and so on. The result is an account upon which an act is virtuous if and only if it results in the highest ranked outcome, vicious if and only if it results in the lowest ranked outcome, and so on—virtue ethics has been consequentialized. Regular virtue ethics, upon which the virtuous action is dictated by the right reason of an agent of good character, and virtue ethics in consequentialized form, upon which the virtuous action brings about the highest ranked, and in this sense best, outcome, yield the same deontic verdicts—they are deontically equivalent. Why not then take virtue ethics to be a form of consequentialism?

The most obvious reason is that we require our moral theories to provide an account of what makes the actions in question morally right, or morally virtuous, or morally prohibited, and the theory in its consequentialized form provides no such explanation. Drawing upon the traditional theory, I determine virtuous and vicious acts through exercising right reason as a person of good character, and it is these judgments of virtuous and vicious actions that determine and explain the ranking of outcomes in consequentialized form. Thus, to put the theory in its consequentialized form I need to presuppose the explanation provided by the nonconsequentialist theory, just as to put Kantian moral theory in consequentialized form I need to presuppose the determination of right and wrong actions through appeal to the unconditioned dictates of pure practical reason. The ranking of outcomes in consequentialized form is for such theories a complete explanatory fifth wheel, presupposing the nonconsequentialist determination of right or virtuous action rather than somehow more perspicuously supplanting it.[15]

Perhaps, however, the point is that because we can place all theories in consequentialized form, this is the proper form for comparing and contrasting them. Yet we have already seen that the consequentialized form of virtue ethics appears to fail to capture the explanatory rationale provided by such a theory. This alone is enough to raise doubts as to

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

whether such theories can be compared without distortion in such a form. But the inadequacy of such a form as a context for comparison is put into high relief with the demonstration that there is a different form, what we can label a "deontologized" form,[16] (p. 32) within which traditionally consequentialist theories can be put alongside traditional deontological and virtue ethical theories without any distortion to their competing rationales. Just as the consequentializing strategy builds the morally relevant features of actions into the relevant ranking of *outcomes* as better, worse, and best, the deontologizing strategy builds the morally relevant features of outcomes into the relevant ranking of *actions* as better, worse, and best. The analogous proposal for "deontologizing" consequentialist theories within the context of such an alternative evaluative framework would have the following form:

> Take whatever features a consequentialist moral theory takes to be relevant to the ranking of outcomes, and build these considerations into the reasons that are relevant to ranking actions as better and worse.

This deontologizing proposal suggests a telic equivalence thesis:

> For any plausible consequentialist theory, we can construct a deontologized version that is equivalent to it.

Deontologizing, thus understood, emphasizes the relationship not between the relevant deontic moral statuses of actions and the ranking and evaluation of *outcomes* as good, better, or best, but between the relevant deontic moral statuses of actions and the telic evaluation of reasons for actions and *actions* themselves as good, better, and, when applicable, best. Just as the consequentializing strategy appropriates any considerations given by nonconsequentialists for determining the deontic moral statuses of actions, building them into the ranking of outcomes, the deontologizing strategy appropriates any considerations given by consequentialists for determining the statuses of outcomes, building them into the reasons that are relevant to ranking actions as better and worse. Thus, for a theory in deontologized form, an act merits the relevant deontic status not if and only if it brings about the *best outcome*, but if and only if it is the *best action*, the action best supported by relevant reasons. Welfare utilitarianism maintains that an action is morally right if and only if it maximizes overall welfare. For the deontologized form of welfare utilitarianism, an action is morally right if it is morally best to do, and it is morally best to do what maximizes overall welfare. For the deontologized form of virtue ethics, an action is virtuous if it is best to do, and it is best to do if it is the result of decision in accordance with the mean as dictated by the right reasoning of agents with good character. For the deontologized form of Kantian ethics, an action is morally right if it is morally best to do, and it is morally best if it is dictated by pure practical reason as respecting the value of persons as ends in themselves. Traditional consequentialist theories are thus those theories, in deontologized form, for which the reasons relevant to the moral ranking of actions reflect only the agent-neutral value of outcomes. Other deontologized theories will hold that the morally relevant reasons reflect the value of persons and wills (Kant), traits of character and relationships (virtue ethics), and so on.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(p. 33) Even granting that Kantian moral theory and Aristotelean virtue ethics can be put in consequentialized form, the point is that such theories cannot be compared in such a consequentialized form without privileging traditionally consequentialist theories that do purport to explain the rightness of actions through appeal to the prior ranking of outcomes as better and worse. It is instead in their deontologized form that all such theories can be compared without distortion. Many consequentialists recognize that their theories are captured without distortion in deontologized form; that is, that theirs are theories upon which the deontic moral status of actions reflects the telic status of actions as better, worse, and best that are distinguished from nonconsequentialist rivals by their accounts of the reasons relevant to the telic ranking of actions.[17] Consequentialist theories appeal to reasons that all reflect value captured without distortion in rankings of outcomes. Rival theories are distinguished by their different accounts of the reasons relevant to the moral ranking of actions, reasons that reflect the fundamental value of persons, wills, relationships, objects, and/or traits of character—things that cannot be captured without distortion within a ranking of outcomes. Deontologized form thus allows, as consequentialized form does not, for comparison without distortion of alternative accounts of reasons and values.

It is sometimes suggested, either directly or by implication, that it is the consequentialist's "compelling idea" that tips the balance in favor of consequentializng rather than deontologizing.[18] The compelling idea is taken to suggest that it is always right, morally required, or at least morally permissible, to do what's best, understood as what promotes the best outcome. Because the consequentialized form relates deontic status of actions to the telic value of outcomes, it is the form that articulates rival theories in a way that is compatible with this compelling idea. But on a plausible alternative interpretation of this idea it would seem to have precisely the opposite implication. The *general* intuitive idea is that it is always right *to do what's best*. It is only by smuggling in precisely the point in dispute, stipulating that what's best is *what promotes the best outcome*, that the consequentializer illicitly appropriates the intuitive appeal of this idea in support of consequentialism. As the availability of the deontologized form demonstrates, however, there is an alternative, more plausible interpretation that tells against traditional consequentialism: not "It is always morally permissible to do *what brings about the best outcome*," but "It is always morally permissible to do *what it is best to do*." This *action* interpretation of the idea seems to capture what is intuitively compelling in the general idea that "It is always morally permissible to do what's best."[19] But the action idea (p. 34) supports deontologizing rather than consequentializing; hence, it provides no support for the consequentializing argument for consequentialism.

It is the deontologizing strategy, then, that seems to capture the intuitive link between the deontic status of actions and goodness, albeit the goodness of actions, and that provides the grid for comparing without distortion alternative theories of the reasons that are relevant to the ranking of actions and the values these reasons reflect. In the face of these obstacles, can anything be said to vindicate the insistence that all plausible alternatives are properly considered in their consequentialized forms—indeed, that they are forms of consequentialism? In fact, the inadequacy of this consequentializing argument

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

for consequentialism clears the way for a deeper argument that the optimal form for any plausible candidate moral theory must be a consequentialized form. This is the consequentialist argument for consequentializing, to which I will now turn.

4. The Consequentialist Argument for Consequentializing

The fact that all nonconsequentialist moral theories can be put in consequentialized form provides no grounds for thinking that they are all forms of consequentialism. If anything, it is the deontologized form that seems to facilitate non-question-begging comparison of rival moral theories based upon the rival accounts that they provide of the reasons relevant to telic evaluation of actions and the values that these reasons reflect. Consequentialist moral theories appear to be merely some among other candidates in deontologized form, those that take the reasons relevant to ranking actions as better and worse all to reflect value captured without distortion in rankings of outcomes.

If, however, the agent-relative consequentializer can provide reasons for holding that it is a condition of the plausibility of any theory of the deontic status of actions that its explanation of this status must be provided through appeal to the prior evaluation of outcomes, then she will provide grounds for thinking that any plausible moral theory must be a form of consequentialism and hence must have a consequentialized form. Such agent-relative consequentializers will reject the agent-neutral commitment, but they will retain—indeed insist upon—the commitment to the explanatory priority of the evaluation of outcomes to the evaluation of actions.

Consequentialists who eschew the agent-neutral commitment do in fact provide an argument for requiring that every plausible candidate moral theory must be a form of consequentialism and hence must have a consequentialized form. The argument in question supports the view that agents ought, in the decisive reasons sense of ought, to perform actions if and only if, and because, they bring about the best outcome relative to the agent. This form of consequentialism does not relate the *moral* ought to agent-neutral rankings of outcomes, but the *rational* ought to rankings of actions as better and worse. (p. 35) Agents rationally ought to perform the action best supported by reasons, and the telic ranking of actions is determined through appeal to reasons that reflect the agent-relative rankings of outcomes. The rationale for the rational deontic evaluation of actions as what the agent ought and ought not to do is provided through appeal to the telic evaluation of outcomes.[20]

This rational form of consequentialism may initially seem every bit as problematic as the more traditional moral forms, running afoul of common sense and alternative theories of what agents rationally ought to do. Intuitively, many reasons to act are not reasons to promote, but to engage in performances that are not promotings. They are reasons to go for a walk, or tell the truth, or contemplate a painting, not reasons to bring about some outcome. If some reasons are not reasons to promote, then the rationales for such reasons

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

will likely not reflect the prior ranking of outcomes; hence, there will be no support for the rational form of consequentialism. I will demonstrate in the remainder of this section that explanatory agent-relative consequentialists can be understood as providing an argument in support of the rational form of consequentialism, this apparent counter-intuitiveness notwithstanding. This argument is grounded in widely held commitments in the theory of practical reasons, the theory of action, and the theory of propositional attitudes. The argument establishes that to accept these default accounts of attitudes and actions is to be committed to the rational form of consequentialism.

Commitment to this rational form of consequentialism follows plausibly from commitment to an outcome-centered account of reasons to act, upon which all reasons to act are reasons to promote outcomes.[21] On such an account good reasons to act are reasons to promote good outcomes, agents have better reasons to promote better outcomes, and agents have the most reason to bring about the best outcome. What agents ought to do, in the standard decisive reasons sense of ought, is bring about the best outcome because it is the best outcome. Michael Smith rightly argues that adoption of such an outcome-centered view of reasons invites a "reduction of one moral concept (the concept of what we ought to do) to another pair of moral concepts (the concepts of goodness and badness)," such that the action that an agent ought to perform will always be one that "produces the most good and the least bad" (2003, 576).

Adoption of this outcome-centered view of reasons supports an outcome-centered view of value. All reasons to act are reasons to promote, and promoting is the rational response to value captured entirely in the ranking of outcomes. This naturally suggests that the rationale for all such reasons is provided by the value of the outcomes to which (p. 36) they are responsive.[22] The best reasons to act reflect the best outcomes, and agents ought to do what is supported by the best reasons. Thus, the deontic evaluation of what agents ought and ought not to do is explained through appeal to the ranking of outcomes as better and worse. Such a rational form of consequentialism in turn provides powerful support for consequentialist moral theory, the outcome-centered view of distinctively *moral* value. If good reasons reflect valuable outcomes, then good distinctively moral reasons will reflect the evaluation of outcomes from a distinctively impartial point of view (as impersonally, agent-neutrally better) or will at least reflect a distinctive and prominent role for valuation from such an impartial standpoint, perhaps interacting with other rankings of outcomes in the determination of reasons that are distinctively moral.

Here an objector might reply that if the case for explanatory agent-relative consequentialism and consequentializing rests on the case for an outcome-centered account of reasons to act, such an argument seems only to push the point of reckoning back a step. After all, we have seen that the outcome-centered account of reasons itself is contrary to intuition: intuitively, not all reasons to act are reasons to promote. Isn't the proper response simply to reject the outcome-centered account of reasons? Yet such an account of reasons finds support in what is arguably the default conception of action, the outcome-centered conception adopted by ethicists ranging from Nagel to Portmore,[23] upon which the end of every action is some outcome to be brought about. If we accept this view that the end of

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

every action is the outcome that it will bring about, and we accept the Anscombean platitude that actions are differentiated by their answers to the "Why?" question, that is, by the reasons that we have for performing them,[24] then it seems clear that every action will be distinguished by the reasons to promote the outcome that is its end, and that every such reason to act will be a reason to promote. The default theory of action provides support for the view that all reasons are reason to promote, and hence for the rational form of consequentialism, and hence for explanatory agent-relative consequentializing of any purportedly plausible candidate rational or moral consequentialist theory.

What if we are tempted to deny such an outcome-centered account of action along with the outcome-centered account of reasons to act? After all, is the end of every action, for example, going for a walk, really some outcome to be promoted? Such a tempting response, however, can appear to be blocked by the standard account of beliefs and desires as propositional attitudes with contrasting directions of fit. The default view (p. 37) holds that actions are rationalized by beliefs and desires. Desires are propositional attitudes with propositional contents; their objects are states of affairs captured by "that clauses." They rationalize actions as bringing about the states of affairs—outcomes—that are their objects.[25] On such an outcome-centered account of desires the actions that they rationalize as promoting the states of affairs that are their objects all do have outcomes as their ends, reasons for action are reasons to promote outcomes, and actions are properly evaluated through appeal to the value of such outcomes. Such outcome-centered accounts of reasons, actions, and attitudes are mutually reinforcing, providing support as a group for the claim that any plausible candidate theory must be a form of consequentialism with a consequentialized form.

Such mutually reinforcing outcome-centered accounts are embedded in rational choice theory and in the standard story of action. The standard form of this standard story takes all actions to be bringings about, in particular to be bringings about of the objects of the practical propositional attitudes that rationalize them, and all such rationalizing practical attitudes are taken to have propositional contents as their objects. The reasons provided by such rationalizing attitudes are reasons to bring about the outcomes that are the objects of the relevant practical attitudes.[26] Similarly, rational choice theory incorporates the outcome-centered views of action, reason, and desire/preference. Actions are rationalized by the agent's preferences, essentially comparative attitudes toward outcomes/options, as bringing it about that such outcomes/options occur.[27] Reasons to act are reasons to bring about preferred outcomes, and the agent has the most reason to bring about the best outcome/option, understood as the outcome that maximally satisfies the agent's preferences among outcomes/options as revealed in the appropriate ranking. Every action is rationalized as bringing about the preferred outcome/option.

Thus, default commitments in the theories of actions and propositional attitudes that are embedded in the standard story of action and in the standard form of rational choice theory provide an argument for the rational form of consequentialism: Agents have most reason to act, hence ought to act, to bring about the best outcome. And if all reasons are reasons to promote, then all moral reasons will be reasons to promote. Any plausible moral

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

theory must be a theory of moral reasons as reasons to promote outcomes. Any commonsense moral reasons that do not appear to be reasons to promote, (p. 38) and any apparently plausible moral theories incorporating such reasons, must, according to this argument, be reinterpreted as fundamentally moral reasons to promote and as theories that incorporate only fundamental moral reasons to promote. That is to say, they must be consequentialized, and their most plausible form must be this consequentialized form, upon which deontic evaluative judgments are rationalized through appeal to antecedent rankings of outcomes. Such plausible candidate theories must have a consequentialized form, because the argument purports to establish that they must be forms of consequentialism. Within the context of this argument, if a candidate moral theory fails to provide a plausible rationale in its consequentialized form, this is grounds not for resisting putting it in this form, but for rejecting the theory as implausible. Appearances notwithstanding, the most plausible form of any apparently nonconsequentialist theory must be a form of consequentialism with a consequentialized form, because the only plausible rationales for the deontic and telic evaluation of actions are provided by appeal to the relevant rankings of outcomes.

Whereas the consequentializing argument for consequentialism relies upon the deontic equivalence thesis, this consequentialist argument for consequentializing involves no commitment to deontic equivalence. It establishes that any plausible theory must be a form of consequentialism. If a theory deontically equivalent to a nonconsequentialist theory has no plausible explanatory rationale, but a variation that has only significant deontic overlap rather than deontic equivalence has a plausible explanatory rationale, it is the latter that will be a plausible alternative form of consequentialism, not the former. Like more traditional agent-neutral consequentializing, such agent-relative consequentializing undertaken in the search for the most plausible explanatory rationale need not aspire to deontic equivalence, and it may well be poorly served by doing so.

5. Against the Consequentialist Argument for Consequentializing

The strongest argument against this consequentialist argument for consequentializing would challenge the prima facie plausibility of these outcome-centered views of actions, reasons, desires, values, and distinctively moral values as a set, transmuting their mutual support into a shared liability. In this section I sketch the outlines of such an argument. This argument unfolds in four steps. First, each of these outcome-centered accounts of actions, attitudes, reasons, and values appears initially to be implausible, and they appear to be implausible in similar ways. Second, there is a standard strategy in each case for deflecting the force of this apparent implausibility, the “degenerate cases” strategy. Third, the apparent plausibility of this strategy is undermined by making explicit a distinction between two senses in which actions, reasons, and attitudes can be understood as involving the bringing about of an outcome, a *rationalizing* sense and a *deflationary* sense. With this distinction clearly in view the apparent plausibility of the degenerate

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(p. 39) cases strategy is exposed as trading upon a conflation between these two senses. Fourth, this distinction between senses of bringing about can be harnessed to demonstrate that other standard arguments for outcome-centered accounts beg the question against opposing, more intuitively plausible alternatives.

Step 1: In each account what appears intuitively to be an outcome-centered subset of the relevant set—of actions, attitudes, and reasons—is put forward as comprising the set in its entirety. Intuitively, only some actions are promotings of outcomes. Others are doings that are seemingly not promotings, for example, going for a walk, contemplating a painting, or telling the truth. Intuitively, only some reasons to act are reasons to promote outcomes, and at least some desires have actions to be performed rather than outcomes to be promoted as their objects. Intuitively, the deontic statuses of only some actions are determined primarily through appeal to the value of outcomes. For example, although I may well hold that promise breakings are bad things to happen and/or that my promise breakings are bad things to happen for me or relative to me, my primary reason to keep my promise typically does not seem to be the promotion of any such outcome, but respect for the persons with whom I interact (or value for my integrity).[28] The obvious question, then, is why we shouldn't simply reject these accounts in light of their apparent systematic implausibility? To reject the accounts is to reject the grounds for the consequentialist argument for consequentializing.

Step 2: A pervasive strategy for attempting to accommodate these problematic cases within outcome-centered accounts treats such seemingly non-outcome-centered cases, whether of attitudes, actions, reasons, or values, as a degenerate subset of outcome-centered cases, those in which the outcome that the agent has reason to promote, and that is the object of the desire, and that is the end of action, is the action itself: Desires to act are desires to bring it about that I act, reasons to act are reasons to bring it about that I act, and the ends of actions that are seemingly not promotings of outcomes are in fact the outcomes that the actions themselves occur. Cases in which I seem to be performing actions of types other than bringings about are on this strategy merely degenerate cases in which the outcome that I am bringing about is (at least in part) an action by me, and this outcome rationalizes the performance of that very action. Following Nagel, my "performance of act B," for example, my keeping my promise, is really "a degenerate case of promoting the occurrence of act B" (1970, 47), for example, of promoting the occurrence of my promise keeping happening. On this view, the seemingly resistant cases are cases in which the end of the action *is* an outcome, but the outcome *is* the action.

This degenerate cases strategy is harnessed to deflect the force of apparent counter-examples in defenses of all of these outcome-centered accounts.[29] Reasons to run are (p. 40) parsed as reasons to bring it about that I run (by running), and the rightness of keeping my promise, even when doing so is neither better for me nor better overall, is parsed as bringing about the best outcome relative to me—that my promise keeping at time *t* happens. Similarly, desires to X are really desires to bring it about that my Xing happens, desires that rationalize actions that bring about outcomes.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Such a strategy might seem to raise as many questions at it answers.[30] But one tempting line of thought recurs in the case for it. The recurrent suggestion is that for any action of phi-ing that does not seem to be rationalized by some outcome that it will bring about, or any reason to phi that does not seem to be a reason to bring about some outcome, or any desire to phi that does not seem to have some outcome to be brought about as its object, it is nonetheless true that by phi-ing, I bring it about that I phi, that is, that my phi-ing happens. Every successful phi-ing by me is in this sense a bringing about of the outcome that my phi-ing occurs, every reason to phi is in this sense a reason to bring it about that I phi, and so on. But if phi-ing is in every case bringing it about that I phi, that the outcome—my phi-ing—happens isn't the outcome that I bring about—"that I phi"—after all the end of my ph-iing, the outcome at which my reason to phi aims, and the real object of my desire to phi? That all such seemingly resistant cases are after all cases of bringing it about that I phi in this sense can seem to suggest that these are after all degenerate outcome-centered cases, cases in which the action is rationalized by an outcome—the outcome that the agent's phi-ing happens.

Step 3: This pervasive strategy for accommodating seemingly resistant cases, however, turns on a conflation of two senses of bringing about. For reasons that are readily apparent, I label the first sense the deflationary sense of bringing about:

> **Deflationary Sense**: In successfully completing an action guided by the reasons she has to undertake it, an agent brings about the outcome that the action occurs.

In acting successfully one brings about the occurrence of one's action (that one's action happens). Bringing about in this sense is a necessary condition of the successful completion of an action, regardless of the reasons one has for undertaking it. This everyday use of bringing about involves no commitment to a rationalizing role for the outcome in question; it merely acknowledges that the occurrence of an action is a necessary condition of the completion of the action in question. Whether the agent's reasons for acting reflect the value of persons, things, character traits, or outcomes, any such reasons to perform an action will, if the action is successfully undertaken for such reasons, bring it about in this sense that the action occurs.

The second is a rationalizing sense of bringing it about:

> **Rationalizing Sense**: An agent's reasons to undertake an action are reasons to promote outcomes; hence, in successfully completing such an action for the (p. 41) reasons the agent has to undertake it, he brings about the outcome that rationalizes its performance.

To characterize my action as such a bringing about is to identify it as a type of action, an action of bringing about an outcome, that is responsive to a type of reasons, reasons provided by the value of outcomes.

With the two senses distinguished, the apparent effectiveness of the degenerate cases strategy is exposed as trading upon an ambiguity between them. For example, that all ac-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

tions are bringings about in this first, deflationary sense does nothing to mitigate the apparent implausibility of treating those that do not seem to be bringings about in the rationalizing sense as instances, albeit degenerate instances, of actions of that type. Only a failure clearly to distinguish these two senses fuels the apparent plausibility of the transition from the fact that every action is in some sense (the deflationary sense) a bringing about to the claim that every action is a bringing about in the rationalizing sense. Nagel's claim that "to act" is "to promote" is nonproblematically true in the merely deflationary sense, since it is a necessary consequent of the performance of any action that the action has been performed. But the claim seems just as nonproblematically false taken in the rationalizing sense. And the truth of the former does nothing to mitigate the apparent substantive falsity of the latter. In the absence of additional arguments, the outcome-centered views of actions, reasons, attitudes, and the evaluation of actions stand as counterintuitive proposals to shoehorn reasons that are not reasons to bring about in the rationalizing sense and actions that are not bringings about in the rationalizing sense and desires that are not desires to bring about in the rationalizing sense into reasons, desires, and actions of this particular subtype. With the distinction clearly in view, we are left with the prima facie implausibility of taking all actions to be bringings about in the rationalizing sense. Because the case for adopting rational and moral forms of consequentialism, hence for consequentializing all candidate moral theories, is built upon these outcome-centered accounts of reasons, actions, and attitudes, the prima facie implausibility of these accounts poses a fundamental challenge to the consequentialist argument for consequentializing.

Step 4: There is, however, no absence of additional arguments for these outcome-centered accounts that ground the case for agent-relative consequentialism and consequentializing (see Hurley 2019). I will briefly take up one such argument, an argument concerning the ends of actions, and show that the distinction between senses of bringing about also provides the tools to undermine this argument. The argument grants that every action is a bringing about in the merely deflationary sense, but maintains also that the end of every action is some outcome to be brought about in the rationalizing sense, hence that every action is rationalized by the value of the outcome that is its end, and that the only plausible interpretation of seemingly resistant cases is the degenerate cases interpretation. The crucial premise of this argument, that the end of every action is the bringing about of some outcome, is often put forward as a virtual platitude. The best argument for such an outcome-centered account of action can seem to be the absence of any plausible alternative.

(p. 42) Yet our distinction between senses of bringing about calls this "platitude" into question. Running brings it about that I have run, telling the truth that I have told the truth, and so on. But with our distinction between senses in place, although it is clear that these are cases of bringing about in the merely deflationary sense, it is not at all clear that they are cases of bringing about in the rationalizing sense, the sense necessary to support the consequentialist argument for consequentializing. The claim that the end of every action is some outcome to be promoted seems simply to beg the question against

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

alternatives to the outcome-centered account and against common sense, according to which actions often are not bringings about of outcomes in the rationalizing sense.

Once the distinction between senses of bringing about opens up the conceptual space for an alternative account of the ends of actions, two commonplace claims about action together suggest a non-question-begging alternative to fill this space. The first is that we *perform* actions—intentional actions are performances.[31] The second is Anscombe's aforementioned point that actions are distinguished by the agent's reasons for undertaking them—by the agent's answer to the "Why?" question. Every intentional action is a performance, and every such performance is distinguished by the agent's reasons for undertaking it. This suggests that the proximate end of every intentional action is the successful completion of the performance in question guided by the agent's reasons for undertaking it (ceteris paribus). The end of every action, every such performance undertaken for reasons, is its completion guided by these very reasons. And to achieve the end of every action, thus understood, is to bring it about, in the deflationary sense, that the action occurs.

If every such performance were a promoting of some outcome in the rationalizing sense, as the outcome-centered account claims, then every action would be guided to completion by the reasons for undertaking it, and the reasons for undertaking it would be reasons for promoting some outcome. Thus, successfully completing the performance would, on such an account, be successfully promoting the outcome. But with this non-question-begging account of the ends of actions in view, upon which each action brings about in the deflationary sense, it becomes clear that such an outcome-centered interpretation of the ends of action is prima facie implausible. I might take myself to have reasons to keep my promise, to complete this performance guided by the reasons for undertaking it, that are provided by the value of the persons with whom I am interacting. I do it because it is a good thing to do, not because my doing it is a good thing to happen. To achieve my end in acting in such a case is to complete the performance of keeping my promise guided by the reasons of respectful interaction with others that rationalize it. These are reasons to perform actions that are not reasons to promote valuable outcomes. This non-question-begging characterization of the ends of actions does not rule out the view that all actions are promotings, but it does leave the view prima facie implausible, and in need of argument. Because the consequentialist argument for consequentializing relies upon such outcome-centered accounts, the argument itself is called into question.

(p. 43) 6. Conclusion

The strategy of consequentializing features that common sense takes to be relevant to the deontic evaluation of actions, building them into the telic evaluation of outcomes, is almost as old as consequentialism itself. The recent rejection by many consequentialists of the commitment to the traditional agent-neutral constraint on ranking outcomes has ushered in consequentializing arguments for consequentialism and consequentialist arguments for consequentializing. While the former fail, the latter ground the case for conse-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

quentializing in deeply entrenched and widely held commitments concerning reasons, actions, and attitudes. These commitments to outcome-centered accounts of reasons, attitudes, and actions dictate that any plausible account of the deontic rational and moral evaluation of actions must be a form of consequentialism and hence that any such plausible candidate must have a consequentialized form. Theories that do not provide a plausible rationale for the evaluation of actions in consequentialized form are, the argument concludes, implausible candidates.

The outcome-centered accounts of reasons, attitudes, and actions that ground the consequentialist argument for consequentializing all run afoul of common sense, and do so in similar ways. One strategy for mitigating this counter-intuitiveness, the degenerate cases strategy, trades upon a conflation of two distinct senses in which we speak of actions as bringing about outcomes. This distinction between senses, in turn, can be harnessed to call into question arguments that draw upon these outcome-centered accounts to dictate that any plausible theory must be a form of consequentialism; hence, that every plausible candidate theory must have a consequentialized form.

The arguments I have sketched here only scratch the surface of the considerations that can be and have been offered in support of and against these outcome-centered accounts, hence both in support of and against the consequentialist argument for consequentializing. I believe that the arguments against will carry the day, but I also take it to be clear that a compelling case has not yet been made either way. Crucially, although the conclusion of this argument for consequentialism and consequentializing is a commitment in normative ethics, the premises supporting this conclusion are not; they are commitments in action theory, the theory of rationality, and the philosophy of mind. Anscombe argued years ago (1958) that we cannot begin to address adequately the question of consequentialism without first addressing such “sub-ethical” questions about the nature of actions and propositional attitudes. The consequentialist argument for consequentializing vindicates her claim, even as it sharpens our understanding of what these subethical questions are.

References

Anderson, Elizabeth. 2001. “Unstrapping the Straightjacket of ‘Preference.’” *Economics and Philosophy* 17:21–38.

Anscombe, G. E. M. 1958. “Modern Moral Philosophy.” *Philosophy* 33:1–18.

(p. 44) Anscombe, G. E. M. 2000. *Intention*. Cambridge, MA: Harvard University Press.

Aristotle. 1999. *Nichomachean Ethics*. Translated by Terence Irwin. Indianapolis: Hackett.

Baumann, Marius. 2019. “Consequentializing and Underdetermination.” *Australasian Journal of Philosophy* 97: 511–527.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Bentham, Jeremy. 1907. *An Introduction to the Principles of Morals and Legislation*. Oxford: Oxford University Press.

Broome, John. 1991. *Weighing Goods*. Oxford: Blackwell.

Brown, Campbell. 2011. “Consequentialize This.” *Ethics* 121:749–771.

Cummiskey, David. 1996. *Kantian Consequentialism*. Oxford: Oxford University Press.

Dreier, Jamie. 1993. “Structures of Normative Theories.” *The Monist* 76:22–40.

Dreier, Jamie. 2011. “In Defense of Consequentializing.” In *Oxford Studies in Normative Ethics Vol. I*, edited by Mark Timmons, 97–119. New York: Oxford University Press.

Frankfurt, Harry. 1978. “The Problem of Action.” *American Philosophical Quarterly* 15:157–162.

Hurley, Paul. 2013. “Consequentializing and Deontologizing: Clogging the Consequentialist Vacuum.” In *Oxford Studies in Normative Ethics, Vol. 3*, edited by ark Timmons, 123–153. Oxford: Oxford University Press.

Hurley, Paul. 2017. “Why Consequentialism’s “Compelling Idea’ Is Not.” *Social Theory and Practice* 43:29–54.

Hurley, Paul. 2018. “Consequentialism and the Standard Story of Action.” *The Journal of Ethics* 22:25–44.

Hurley, Paul. 2019. “Exiting the Consequentialist Circle: Two Senses of Bringing It About.” *Analytic Philosophy* 60:130–163.

Louise, Jennie. 2004. “Relativity of Value and the Consequentialist Umbrella.” *Philosophical Quarterly* 54:518–536.

Mill, J. S. 2001. *Utilitarianism*. Indianapolis: Hackett.

Moran, Richard, and Stone, Martin. 2011. “Anscombe on Expression of Intention: An Exegesis.” In *Essays on Anscombe’s Intention*, edited by A. Ford, J. Hornsby, and F. Stoutland, 33–75. Cambridge, MA: Harvard University Press.

Nagel, Thomas. 1970. *The Possibility of Altruism*. Oxford: Clarendon Press.

Nye, Howard, Plunkett, David, and Ku, John. 2015. “Non-Consequentialism Demystified.” *Philosophers’ Imprint* 15:1–28.

Parfit, Derek. 2011. *On What Matters, Vol. I*. Oxford: Oxford University Press.

Peterson, Martin. 2010. “A Royal Road to Consequentialism?” *Ethical Theory and Moral Practice* 13:153–169.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Portmore, Douglas. 2007. “Consequentializing Moral Theories.” *Pacific Philosophical Quarterly* 88:39–73.

Portmore, Douglas. 2011. *Commonsense Consequentialism*. Oxford: Oxford University Press.

Portmore, Douglas. 2019. *Opting for the Best*. New York: Oxford University Press.

Sachs, Benjamin. 2010. “Consequentialism’s Double-Edged Sword.” *Utilitas* 22:258–271.

Sauer, Hanno. 2019. “The Cost of Consequentialization.” *Metaphilosophy* 50:100–109.

Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, MA: Belknap Press, Harvard University Press.

Scheffler, Samuel. 1982. *The Rejection of Consequentialism*. Oxford: Oxford University Press.

Schroeder, Andrew. 2017. “Consequentializing and Its Consequences.” *Philosophical Studies* 174:1475–1497.

Schroeder, Mark. 2007. “Teleology, Agent-Relative Value, and “Good.’” *Ethics* 117:265–295.

Sen, Amartya. 1973. “Behaviour and the Concept of Preference.” *Economica* XL:241–59.

(p. 45) Sen, Amartya. 1983. “Evaluator Relativity and Consequential Evaluation.” *Philosophy and Public Affairs* 12:113–132.

Smith, Michael. 2003. “Neutral and Relative Value after Moore.” *Ethics* 113:576–598.

Smith, Michael. 2009. “Two Kinds of Consequentialism.” *Philosophical Issues* 19:257–272.

Smith, Michael. 2012. “Four Objections to the Standard Story of Action (and Four Replies).” *Philosophical Issues* 22:387–401.

Sumner, Wayne. 1996. *Welfare, Happiness, and Ethics*. Oxford: Oxford University Press.

Tenenbaum, Sergio. 2014. “The Perils of Earnest Consequentializing.” *Philosophy and Phenomenological Research* LXXXVIII:233–240.

Thompson, Michael. 2008. *Life and Action*. Cambridge, MA: Harvard University Press.

Wiggins, David. 2006. *Ethics*. Cambridge, MA: Harvard University Press.

Notes:

([1]) Advocates of agent-relative (evaluator-relative) consequentialism include Amartya Sen (1983), Jamie Dreier (1993), Michael Smith (2003), Douglas Portmore (2011), Jennie Louise (2004), and Martin Peterson (2010).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([2]) There are certain respects in which this account of consequentialized form is too restrictive, that is, in which a moral theory can, strictly speaking, be a form of consequentialism without having in this sense a consequentialized form. Such overly restrictive aspects do not affect the arguments to follow; hence, I set them aside. But see Portmore's maximalist consequentialism (2019, 156–161 and 196) for an example of a form of consequentialism that does not have such a consequentialized form.

([3]) This characterization of forms of consequentialism is narrower than some and broader than others. One respect in which it is broader is that it encompasses forms of consequentialism that explain the deontic evaluations of actions as what ought and ought not to happen, in the standard decisive reasons sense of ought, through appeal to the relevant telic evaluation of outcomes. That is to say, it comprehends rational in addition to distinctively moral forms of consequentialism.

([4]) See, for example, Peter Railton's (1988) rejection of monism in favor of pluralism.

([5]) See David Cummiskey (1996) for such a proposal to consequentialize Kantian respect for persons within an agent-neutral standpoint, and Samuel Scheffler (1982, chap. 4) for a discussion of the proposal to consequentialize rights.

([6]) Although the rejection of agent-neutral rankings of outcomes in favor of an agent-relative alternative is the most significant development in this new wave of consequentialist views, rejections of the time, place, and world neutrality of rankings in favor of time-, place-, and world-relative alternatives have also been proposed. For an insightful discussion of these other dimensions of the neutrality and relativity of rankings, see Hammerton's chapter (Chapter 3) in this volume.

([7]) The best outcome *relative to the agent* can also diverge from the prudentially best outcome *for the agent*. Thus, it may be better *for the agent* to break her promise (e.g., it will save her considerable hardship) and better overall to break her promise (e.g., it will prevent more promises overall from being broken), but nonetheless be better *relative to her* to keep her promise (e.g., it will maximize her keeping of her promises).

([8]) See Michael Smith (2009), Jamie Dreier (2011), and Douglas Portmore (2011, chap. 4) for demonstrations that agent-relative consequentialism can accommodate deontic constraints without indirection. Mark Schroeder challenges the very meaningfulness of such claims concerning what is best relative to the agent as opposed to what is best for the agent and what is best overall (2007). I note this important challenge here only to set it aside in what follows.

([9]) Such consequentializers who appear to eschew the explanatory commitment include Jamie Dreier (2011, sec. 4), Martin Peterson (2010), and Jennie Louise (2004).

([10]) For an insightful discussion of this tendency to run together different and in some cases inconsistent arguments involving consequentializing, see Andrew Schroeder (2017).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([11]) For discussion and criticism of such pragmatic arguments, see again Schroeder (2017, 1488–1495).

([12]) See Portmore (2007, 40) and Dreier (2011, 97).

([13]) For such a challenge, see Campbell Brown (2011).

([14]) See, for example, Jennie Louise's claim that because all normative theories are "describable as consequentialist," the question is no longer "whether we should be consequentialists or not" (2004, 536), and Martin Peterson's claim that the ability to put such theories in consequentialized form shows that "people advocating rival moral theories just make slightly different claims about how to evaluate consequences" (2010, 155).

([15]) Portmore discusses the shortcomings of such "Footian" consequentializing (2011, 113). Arguments stressing the explanatory impoverishment of many consequentialized forms of nonconsequentialist theories are also provided by Andrew Schroeder (2017, 1478–1482), Sergio Tenenbaum (2014), Marius Bauman (2019), and Hanno Sauer (2019).

([16]) See my 2013 for the development of this deontologizing alternative, and Benjamin Sachs' related argument that "the truth of Consequentializability would also, surprisingly, yield the result that we can construct more plausible versions of *non*-consequentialism" (2010, 262). Jamie Dreier also allows that all plausible theories can be put in deontologized as well as consequentialized form but makes a case for preferring their consequentialized form (2011, sec. 4).

([17]) Thus, John Broome maintains that "the rightness of actions is determined by their goodness" (1991, 6), and Douglas Portmore maintains that "an agent objectively ought to perform some particular action if and only if it is, in fact, the best alternative," where the best alternative action is "the alternative that she has the most reason to perform" (2011, 12).

([18]) See, for example, Dreier (2011, 115).

([19]) I my 2017, I develop this distinction among the General Idea, the Outcome Idea, and the Action Idea in more detail. In particular, I show (2017, 36–38) that it is the Action Idea that seems better to capture the intuitive force of the General Idea, because it captures the deep intuitions that if we have the best reasons for performing some course of action (it is the best course of action), we ought, rationally, to perform it, and that it is never morally impermissible to do what we ought, rationally, to do. For a similar argument, see David Wiggins (2006, 215–218).

([20]) Michael Smith explicitly defends what I here characterize as a rational form of consequentialism, according to which agents ought to perform the action that "produces the most good and the least bad" (2003, 576). We have seen that Portmore similarly holds that "an agent objectively ought to perform some particular action if and only if it is, in

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

fact, the best alternative," where the best alternative is the alternative that brings about the highest ranked outcome (2011, 12).

([21]) Such an outcome-centered account of reasons to act is commonly referred to as a teleological conception of reasons. See Scanlon for an argument against the teleological conception (1998, 78–87) and Portmore for a defense of such a conception (2011, chap. 3).

([22]) Such an account need not deny that things other than outcomes can have intrinsic value; it need only maintain that their relevance to reasons for action is captured without distortion in a ranking of better and worse outcomes.

([23]) Thus, Thomas Nagel suggests that even the performance of an action that does not appear to be a bringing about, for example, the "performance of act B," is really "a degenerate case of promoting the occurrence of act B" (1970, 47), hence that all actions are promotings of outcomes. See also Portmore's assertion that all actions not only "alter the way the world goes"; they all "aim at making the world go a certain way" (2011, 56), that is, at the promotion of some outcome.

([24]) In particular, she claims that an intentional action is an action "to which a certain sense of the question "Why?" is given application; the sense … in which the answer, if positive, gives a reason for acting" (2000, 9).

([25]) For characterizations of this default outcome-centered account of desire, see Michael Smith (2012, 387) and Derek Parfit (2011, 43).

([26]) See Harry Frankfurt (1978) for a presentation and criticism of the standard story, and Michael Smith (2012) for a presentation and defense. In my (2018) I explore this relationship between moral consequentialism and the standard story in more detail.

([27]) For discussions of this role for preferences, see Elizabeth Anderson (2001, 22–23) and Amartya Sen (1973). An alternative understanding of rational choice theory denies that preferences play this explanatory, rationalizing role, limiting them instead to a predictive role. Preferences, on such a merely predictive model, do not purport to explain choice; they reflect choice, and such modeling of choices in rankings of outcomes is defended on pragmatic grounds (see Schroeder 2017), for example, as allowing the deployment of formal tools that facilitate prediction. Such predictive models require no commitment to outcome-centered views of action, reason, and desire/preference.

([28]) Howard Nye, David Plunkett, and John Ku highlight and defend this intuitive distinction between our motives "that are state-directed, or motives to bring about certain states of affairs" and "motives that are act-directed… motives simply to do certain things" (2015, 5).

([29]) Particularly clear examples are to be found in Douglas Portmore's account of reasons (2011, chap. 3), Wayne Sumner's account of desires (1996, 124), and Jamie Dreier's account of the constitutive consequences of actions (2011, 99–100).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([30]) For a discussion of these additional questions raised by the degenerate cases strategy see my 2019, 10–12.

([31]) For accounts that emphasize the nature of actions as performances, see Michael Thompson (2008, 127–138) and Richard Moran and Martin Stone (2011).

Paul Hurley

Paul Hurley is the Sexton Professor of Philosophy at Claremont McKenna College, The Claremont Colleges. His research focuses primarily upon ethics, particularly the debate between consequentialists and their critics, but he has also published in metaethics, action theory, and the history of ethics. He is the author of Beyond Consequentalism (Oxford University Press, 2009) and of over two dozen articles. His current project is to demonstrate that the central arguments for consequentialism are grounded below ethics, in outcome-centered theories of actions and attitudes, and to challenge the case for consequentialism at this deeper level.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Fault Lines in Ethical Theory

Shyam Nair

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.24

Abstract and Keywords

The verdicts that standard consequentialism gives about what we are obligated to do crucially depend on what theory of value the consequentialist accepts. This makes it hard to say what separates standard consequentialist theories from nonconsequentialist theories. This article discusses how we can draw sharp lines separating standard consequentialist theories from other theories and what assumptions about goodness we must make in order to draw these lines. The discussion touches on cases of deontic constraints, cases of deontic options, and cases involved in the so-called actualism/possibilism debate. What emerges is that there are various interesting patterns relating the different commitments of consequentialism, different principles about obligation and about goodness, and different rules concerning how facts about values determine facts about obligation.

Keywords: constraints, options, actualism, possibilism, consequentializing, axiology, deontic logic, decision rules

1. Introduction

WHEN we introduce students to moral philosophy, we draw sharp contrasts between different ethical theories. Chief among these contrasts is the contrast between consequentialist and nonconsequentialist moral theories. We often illustrate the contrast with cases like the Williams's integrity challenges (1973) and Foot's Trolley Problem (1978).[Thanks to Doug Portmore for comments on this article.]

But the contemporary literature about moral philosophy does not fit well with this tidy picture. As the articles in this handbook attest, consequentialism is a label for a startling variety of different theories that can often make different predictions about these cases. Indeed, some have gone so far as to argue that consequentialism is such an accommodating framework that it can deliver the same verdicts as any plausible nonconsequentialist theory.[2] This is because, these theorists argue, there is always some ranking of outcomes according to their goodness that can deliver whatever result about a case that we wish.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

This article investigates to what extent we can draw lines that separate standard consequentialist theories from other theories. In my usage, a *fault line* is a case or really a structural description of a class of cases that can be used to separate moral theories because the theories give different verdicts about the class of cases. I will be interested in what fault lines there are and in the closely related question of how to establish that these fault lines exist.

(p. 68) We begin by exploring two of the most studied kinds of cases that are supposed to separate consequentialist theories from deontological theories, cases of deontic constraints and cases of deontic options (section 2).[3] We then discuss another family of fault lines that can be described using certain simple abstract principles (section 3). These fault lines separate standard consequentialist theories from certain nonconsequentialist theories that nonetheless give the value of outcomes an important place in determining the deontic status of acts (section 4). Finally, we close by commenting on how our discussion bears on broader issues in ethical theory (section 5).

2. Deontic Constraints and Options

Cases of deontic constraints come in a variety of forms.[4] Here is a simple case:

> Mac faces a choice of whether or not to murder an innocent person, Jack. If Mac does not murder Jack, Frank and Hank will each murder a different innocent person. If Mac does murder Jack, he will prevent Frank and Hank from murdering.

Many believe that it would be wrong for Mac to murder Jack, even though Mac's murdering Jack would prevent Frank and Hank from murdering two innocent people. Many nonconsequentialist theories give this result about the case. On the other hand, it is hard to see how *standard act consequentialism*—S is obligated to do *x* iff the outcome of S's doing *x* is better than the outcome of S's refraining from doing *x*—can get this result.[5] Given that the outcome of Mac's murdering Jack contains just one innocent person being murdered and given that the outcome of Mac's refraining from murdering Jack contains two innocent murders, it seems that the outcome of Mac's murdering Jack is better than the outcome of Mac's refraining from murdering Jack. Thus, standard act consequentialism appears to be committed to the idea that Mac is obligated to murder.

And cases like Mac's are just one of a class of structurally similar cases that raise similar problems such as the following kinds of cases:

- cases where an agent can break a promise in order to prevent others from breaking their promises
- cases where an agent can forgo providing a benefit for one of her nearest and dearest in order to prevent others from forgoing providing a benefit for one of their nearest and dearest

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(p. 69) One standard gloss on the structural similarity between these cases is that they are cases where an agent faces a choice of whether to perform an act where performing that act will prevent others from performing an act of the same morally relevant type. According to some nonconsequentialist theories, there are cases where it is wrong for an agent to perform an act (e.g., murder, break a promise) even though doing so would prevent others from performing an act of the same morally relevant type (e.g., murder, break a promise). But according to standard consequentialism, it appears there can be no such cases.

Cases of deontic options, like cases of deontic constraints, come in a variety of forms. Here is a simple case of this sort:

> Carol works for a nonprofit helping the global poor. She works directly with those affected by poverty and is passionate about it. She could quit her job and work in finance and donate all her money to charity. She would help slightly more people by donating money rather than working directly with the global poor.

Many people believe that it is permissible for Carol to continue working for the nonprofit and permissible for her to work in finance. Standard consequentialism, on the other hand, appears to entail that Carol is required to quit her job and work in finance because this will help slightly more people. Many nonconsequentialist theories recognize a special dispensation permitting an agent to pursue the projects that matter the most to her, but standard consequentialism is accused of being incompatible with this.

2.1. Dimensions of Importance

Suppose that it has been shown that consequentialist and nonconsequentialist theories are committed to giving different results about cases of deontic constraints and cases of deontic options. These cases then would be fault lines that separate these theories. Can anything more be said about these fault lines?

One way to think about this question is to ask whether the fault lines target only some of the core commitments of consequentialism or all of them. There are at least three core commitments of consequentialism. The first two are the commitments we get from separating the biconditional that is the statement of consequentialism into a pair of conditionals:

> *Left-to-Right*: if S is obligated to do *x*, then the outcome of S's doing *x* is better than the outcome of S's failing to do *x*
>
> *Right-to-Left*: if the outcome of S's doing *x* is better than the outcome of S's failing to do *x*, then S is obligated to do *x*

The third commitment is an entailment of consequentialism that is sometimes called the "compelling idea," the idea that it is always permissible to do what is best:

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Compelling Idea: if the outcome of S's doing *x* is the best outcome, it is permissible for S to do *x*

(p. 70) We can now ask the more specific question of what cases of constraints and options say about each of these theses.

Begin with cases of constraints. If it is true that the outcome of Mac's murder is better than the outcome of Mac's refraining from murder and it is true that Mac is not obligated to murder and that Mac is obligated to refrain from murder, then cases of constraints are incompatible with all three theses. To see that it is incompatible with *Left-to-Right*, consider that Mac is obligated to refrain from murdering even though the outcome of murdering is better than the outcome of refraining from murdering. To see that it is incompatible with *Right-to-Left* and *Compelling Idea*, consider that the outcome of Mac murdering is the best outcome available to Mac and yet it is not permissible for Mac to murder.

On the other hand, cases of deontic options are only incompatible with *Right-to-Left*. In cases of deontic option it is merely permissible (so not obligatory) for Carol to work in finance, even though the outcome of her doing so is better than the outcome of her not doing so (and dedicating her life to working at the nonprofit). On the other hand, these cases are no threat to *Left-to-Right* because no act is obligatory in this case. And these cases pose no threat to *Compelling Idea* because the act with the best outcome is permissible.

Thus, there is a sense in which cases of deontic constraints are more *deeply* in tension with the consequentialist framework than cases of deontic options. I will, then, say we can analyze putative fault lines according to how deeply incompatible with consequentialism they are where one measure of depth is which commitments of consequentialism must be rejected.

Of course, unlike the ordinary notion of depth, this notion cannot be numerically measured. Instead, it is better understood as a useful indicator. So though we may not be able to so precisely compare the depth of fault lines, we can still use adjectives related to depth as helpful labels that tell us to look more closely at exactly which commitments are rejected by each fault line.

2.2. Assumptions about Goodness

In order to establish that consequentialism is incompatible with cases of deontic constraints and cases of deontic options, we need to make substantive assumptions about the goodness of outcomes, for example, that the outcome containing two murders is worse than the outcome containing one.

And these assumptions are initially quite plausible. But what is to prevent the consequentialist from rejecting these assumptions in favor of a different account of goodness? For example, what is to prevent the consequentialist from claiming that keeping one's hands clean from murder is better than preventing those who already have murderous consciences from murdering?

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

More generally, consequentialists who are pluralists about the good have considerable resources to accommodate putative fault lines. Campbell Brown observes that to those who claim that a certain case is incompatible with consequentialism:

(p. 71)

> the consequentialist [… may] reply. “Your story presupposes a certain account of what makes consequences better or worse, a certain ‘theory of the good,’ as we consequentialists like to say. Consequentialism, however, is not wedded to any such theory. We already knew that combining consequentialism with some theories of the good would have implausible results; that’s what utilitarianism has taught us. In order to reconcile consequentialism with the view that this action you’ve described is wrong, we need only to find an appropriate theory of the good, one according to which the consequences of this action would not be best. You say you’re concerned about the guy’s rights? No worries; we’ll just build that into your theory of the good. Then you can be a consequentialist too.” (2011, 749–750)

Of course, there may be arguments against the theory of goodness that consequentialism requires to get the desired results in these cases. But there is no easy road from simple verdicts about cases to the rejection of consequentialism. Additional argument concerning what is good is needed.

If the strategy that Brown outlines works, it would show that we cannot establish the existence of fault lines separating consequentialist from nonconsequentialist theories while being *neutral* about what theory of goodness is correct. I call the question of what it takes to establish a fault line the question of on *what grounds* the fault line can be established. And what we have just seen is that there is an interesting question of whether any fault line can be established on *neutral* grounds.

This gives us a second dimension of evaluation by which to consider various fault lines.

2.3. There Are No Deep Fault Lines on Neutral Grounds

We have already seen that given certain assumptions about goodness, we can establish that cases of deontic constraints form a deep fault line. But we have also seen that these assumptions are substantive. In this subsection, we will see that though some fault lines *can* be established on neutral grounds, no *deep* fault line can be established on *neutral* grounds.

2.3.1. A Fault Line on Neutral Grounds

We begin by seeing that we can establish that cases of deontic constraints are incompatible with consequentialism on *neutral* grounds. The only assumptions about goodness that we will make are standard logical assumptions about the betterness relation (e.g., that it is irreflexive, transitive, and complete).[6]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Consider a case where if Mary doesn't murder, John will and where if John doesn't murder, Mary will. Intuitively, this is a case of constraints where it is wrong for Mary to murder and wrong for John to murder.

(p. 72) There are only two possible ways things could go:

$\mathbf{w}_1$: Mary murders, John doesn't

$\mathbf{w}_2$: Mary doesn't murder, John does

No matter how these outcomes are ordered with respect to betterness, we must reject either the claim that it is obligatory for Mary to not murder or the claim that it is obligatory for John to not murder (and not obligatory for them to murder). For suppose $\mathbf{w}_1$ is better than $\mathbf{w}_2$, in this case consequentialism entails that it is obligatory for Mary to murder. Analogously, if $\mathbf{w}_2$ is better than $\mathbf{w}_1$, it follows that it is obligatory for John to murder. Finally, if $\mathbf{w}_1$ and $\mathbf{w}_2$ are equally good, it follows that it is not obligatory for either of them to refrain from murder. Thus, it follows that no matter what theory of goodness we accept, consequentialism is incompatible with cases of deontic constraints.

2.3.2. Impossibility and Modesty

That said, this argument does not establish that there is a *deep* fault line for it is compatible with *Left-to-Right* and *Compelling Idea*. If $\mathbf{w}_1$ and $\mathbf{w}_2$ are equally ranked, *Left-to-Right* must hold because no act is obligatory in such a case, and *Compelling Idea* must hold because all acts are permissible in such a case. More generally, in order to show that either of these theses fails, one outcome must be ranked ahead of another. This shows that no neutral argument can ever establish a deep fault line because there is nothing about the logical structure of goodness that entails that one outcome is strictly better than another.

This teaches us that in order to establish deep fault lines, we will need to appeal to some substantive principle that tells us one outcome is better than another. But these principles need not be as strong as the claim that violating rights is bad or donating to charity is good. Instead, we can rely on *modest* theoretical principles.

One such principle is a Pareto-like principle that I will call *Unanimity*:

> *Unanimity* if every agent ought to prefer an outcome $\mathbf{w_i}$ to another outcome $\mathbf{w_j}$, then $\mathbf{w_i}$ is better than $\mathbf{w_j}$

This is a high-level minimal principle connecting the preferences that every agent ought[7] to have with goodness. Though I do not believe that there is any snappy argument in favor of this principle, it is quite plausible.

In other work (Nair 2014), I argue that if we accept this principle, we can show that a certain class of cases of deontic constraints forms a deep fault line. Since the principle is neither a purely logical principle nor a more substantive claim about goodness, I say we can establish that these cases form a deep fault line on relatively *modest* (albeit not neutral) grounds. The argument for this is too lengthy to be rehearsed (p. 73) here.[8] And the de-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

tails are in any case inessential. But what matters for our purposes is that we may uncover certain modest principles about value that can be used to establish fault lines in ethical theory. The next section adopts this approach.

3. Structural Descriptions of Cases and Standard Consequentialism

As we have seen, typically fault lines are paradigmatic examples (e.g., cases where common-sense morality suggests it is wrong to murder even to prevent more murders) or specified with informal glosses (e.g., cases where it is wrong to do an act even though that act would prevent the performance of more acts of the same morally relevant type). But little work has been done to describe these fault lines formally.

This is understandable as these cases have quite complex normative and causal structures. Nonetheless, in this section, we will consider how simple formal principles can specify fault lines.

3.1. The Logical Structure of Cases

We focus on extremely simple cases involving only what an agent at a time is obligated to do (though these acts may occur at distinct times). We will use O as an obligation operator and take it to be implicitly indexed to an agent at a time. We will also assume that O officially is a propositional operator rather than an operator on acts themselves. But often, when speaking informally, we will freely switch between treating the prejacents of O as propositions or as acts. In this setting, we can write down some simple sentences and take them to characterize a class of cases. For example, presumably the sentence "O(A) or ~O(A)" characterizes every case. And presumably every theory says "yes" to the existence of this class and "no" to the existence to a nonempty complement of this class. So it does not form a fault line.

But there are other more interesting cases as well. Here is one example:

> *Agglomeration* O(A) & O(B) → O(A & B)

Agglomeration is a structural description of a class of cases in which either O(A) & O(B) is false or O(A & B) is true. The complement of this class is the cases where O(A) & O(B) is true and O(A & B) is false. For *Agglomeration* to specify a fault line would be for there (p. 74) to be a dispute among theories about whether the class specified by it or by its complement is nonempty. As we will see, there is such a dispute about whether the complement of this class of case is nonempty.

In a similar vein, we can ask about the class of cases and complement of the class of cases characterized by the following principles:

> *Inheritance* If A entails B, then O(A) entails O(B)

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

No Conflicts If O(A_1) is true, O(A_2) is true, … , and O(A_n) is true, then {A_1, A_2, … , A_n} is consistent

No Strict Conflicts ~(O(A) & O(~A))

Later, we will discuss the plausibility of these principles (section 4.5). But for now, we turn to an initial discussion of how consequentialism as standardly formulated is incompatible with these principles.

3.2. Standard Consequentialism and the Ubiquity of Fault Lines

As discussed earlier, standard act consequentialism is the following claim:

S is obligated to do *x* iff the outcome of S's doing *x* is better than the outcome of S's failing to do *x*

where we say for a possible world, **w**:

w is the outcome of S's doing(/refraining from doing) *x* iff if S were to do(/refrain from doing) *x*, then **w** would obtain[9]

We will look in greater detail at the features of this formulation in section 4.1. But for now, what I wish to point out is that *Agglomeration*, *Inheritance*, and *No Conflicts* appear to be fault lines in that standard consequentialism accepts the existence of cases that falsify these principles while other theories that accept these principles are committed to there being no such cases.

Many cases in the literature have been offered that attest to standard consequentialism's rejection of these principles.[10] Here we can consider one due to Michael Zimmerman:

(p. 75)

I have been invited to attend a wedding. The bride-to-be is a former girlfriend of mine; it was she who did the dumping. Everyone, including me in my better moments, recognizes that she was quite right to end our relationship; we were not well suited for one another, and the prospects were bleak. Her present situation is very different; she and her fiancé sparkle in one another's company, spreading joy wherever they go. This irks me to no end, and I tend to behave badly whenever I see them together. I ought not to misbehave, of course, and I know this; I could easily do otherwise, but I do not. The wedding will be an opportunity for me to put this boorishness behind me, to grow up and move on. The best thing for me to do would be to accept the invitation, show up on the day in question, and behave myself. The worst thing would be to show up and misbehave; better would be to decline the invitation and not show up at all. (2006, 153)

Zimmerman adds to fill out the case "if I accepted the invitation, I would show up and misbehave (whereas I would not do this if I declined). I need not misbehave (for, as noted, I could easily do otherwise); nonetheless, this is what I would in fact do."

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Here Zimmerman can accept the invitation in two ways. One way would be to accept the invitation and go on to behave well; the other to accept the invitation and go on to behave poorly. If he were to accept the invitation, he would go on to behave poorly. He can, on the other hand, decline the invitation. The best outcome is the outcome in which Zimmerman accepts and behaves well. The middle outcome is the outcome in which Zimmerman declines the invitation. The worst outcome is the outcome in which Zimmerman accepts and behaves poorly. The outcome of accepting is the outcome in which Zimmerman accepts and behaves poorly.

According to standard consequentialism, Zimmerman is obligated to decline because the outcome of Zimmerman declining is the middle outcome, which is better than the outcome of Zimmerman not declining (i.e., accepting), which is the worst outcome. By similar reasoning, it is not the case that Zimmerman is obligated to accept. But Zimmerman is obligated to accept and behave well because the outcome of this is best.

If we let "*Accept*" express the proposition that Zimmerman accepts the invitation, "*Behave Well*" express the proposition that Zimmerman behaves well, and "*Decline*" express the proposition that Zimmerman declines the invitation, we have O(*Accept & Behave Well)*, O(*Decline*), ~O(*Accept*). This falsifies *Inheritance*. Next, since one cannot accept, behave well, and decline, this also falsifies *No Conflicts*. Finally, we assume throughout an impossible claim is never obligatory.[11] So we have ~O(*Accept & Behave Well & Decline)*. This, then, falsifies *Agglomeration* as well. Thus, these principles appear to represent a fault lines in ethical theory in that standard consequentialism cannot accept them and other theories do accept them.[12]

But do examples like Zimmerman's rest on substantive assumptions about goodness?

(p. 76) 3.3. Nonneutral Grounds

To check whether they do, let's begin by understanding the structure of Zimmerman's example: There is an act *A* (e.g., *Accept*) that can be done in two incompatible ways, *A & B* (e.g., *Accept & Behave Well*) and *A & ~B* (e.g., *Accept & ~Behave Well*).[13] These acts result in distinct incompatible outcomes, $\mathbf{w}_{(A\ \&\ B)}$ and $\mathbf{w}_{(A\ \&\ \sim B)}$. Further, we assumed that there is another act *C* (e.g., *Decline*) that has a third distinct outcome, $\mathbf{w}_C$. Finally, we claimed that if one were to do *A* (e.g., *Accept*), one would do it some particular way such as *A & ~B* (e.g., *Accept & ~Behave Well*). Evidently, these claims make no assumptions whatsoever about the goodness or badness of outcomes. And they are, in any case, thoroughly innocuous.

But the case relied on some further assumptions. In particular, I claimed that we can rank the outcomes so that $\mathbf{w}_C$ is strictly in the middle (i.e., it is strictly better than exactly one of $\mathbf{w}_{(A\ \&\ B)}$ and $\mathbf{w}_{(A\ \&\ \sim B)}$ and strictly worse than exactly one of $\mathbf{w}_{(A\ \&\ B)}$ and $\mathbf{w}_{(A\ \&\ \sim B)}$). In the example, I claimed the outcome of declining the invitation is strictly better than the outcome of accepting and behaving poorly and strictly worse than the outcome of accepting and behaving well.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

But we already know for the reasons given in section 2.3.2 that these assumptions about goodness must be nonneutral: The assumptions entail that some outcome is ranked ahead of another outcome. But there is nothing about the minimal structural properties of goodness that would tell us this. So the argument that standard consequentialism must reject the aforementioned principles relies on some substantive assumptions about goodness.

3.4. Modest Grounds

But what are the weakest assumptions that we actually need to make about goodness in order to establish this fault line? Or to frame this issue in a different way, is there a theory of goodness that we can supplement standard consequentialism with so as to show that it can accept *Inheritance*, *Agglomeration*, and *No Conflicts*?

We already saw that in Zimmerman's example the act of accepting can be done in two ways: one can accept and go on to behave well or accept and go on to behave poorly. And in Zimmerman's example there was also the act of declining which is incompatible with all of these acts. The crucial additional value assumption that is made in Zimmerman's example is that the value of the outcome of declining can be strictly in between the value of these two different ways of accepting. It assumes that we can in fact pry apart the value of two outcomes in which one accepts.

This suggests that if we adopt a theory of goodness that does not allow us to pry apart the value of two outcomes in which an agent accepts, Zimmerman's case would not be (p. 77) enough to establish that consequentialism is incompatible with *Inheritance, Agglomeration*, and *No Conflicts*. More generally, it suggests the conjecture that if value of the outcomes in which a given act, *A*, occurs cannot be "splintered" in the sense that there is an incompatible act *B* whose outcome is strictly in two distinct outcomes in which *A* occurs, then standard consequentialism can accept *Inheritance*, *Agglomeration*, and *No Conflicts*. This conjecture can be formalized as follows:

> *Value Non-Splintering* if *A* is true at $\mathbf{w}_i$ and at $\mathbf{w_j}$, then for any $\mathbf{w}_k$ where *A* is false, if $\mathbf{w}_i$ is strictly better than (/worse than/equally good as) $\mathbf{w}_k$, then $\mathbf{w}_j$ is better than (/worse than/equally good as) $\mathbf{w}_k$

where *A* is an act statement that the agent does some act (e.g., S does *x*) and where $\mathbf{w}_i$, $\mathbf{w}_j$, $\mathbf{w}_k$ are outcomes.[14]According to this principle, it cannot be that the outcome of *Decline* is strictly in between the outcomes of *Accept and Behave Well* and the outcome of *Accept and ~Behave Well*. This is because the outcome of *Accept and Behave Well* and the outcome of *Accept and ~Behave Well* are both outcomes in which *Accept* holds. And the principle tells us that no outcome where *Accept* does not hold can come strictly between two outcomes in which *Accept* holds. Thus, if this principle holds, we cannot use Zimmerman's case to establish that consequentialism is incompatible with *Inheritance*, *Agglomeration*, and *No Conflicts*.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Of course, it is unsurprising that we get these results in Zimmerman-type cases as the principle is tailored to handle that case. But our more general conjecture is also true: Standard consequentialism paired with *Value Non-Splintering* entails *Inheritance*, *Agglomeration*, and *No Conflicts*. Claims 1–3 in the appendix provide the relevant proofs of the conjecture.

Value Non-Splintering and its negation are nontrivial claims. That said, they are quite different from the claim, for example, that protection of rights makes no contribution to goodness. One obvious difference is its generality.

Another difference is that *Value Non-Splintering* is quite implausible in its own right. One way to think about *Value Non-Splintering* is that it says that for a given act *A* the value of *A* has a kind of lexical priority in ordering outcomes in which *A* with respect outcomes in which *~A*. Outcomes in which *A* may be better or worse than or equal to one another. But when it comes to comparisons to outcomes in which *~A*, the presence of *A* alone suffices to settle the ranking of outcomes. While this property may be sensible for certain acts that are especially morally awful or especially morally good, it is not sensible for every act. Consider ordinary acts such as the act of deciding to spend time reading a book, going to the movies, or eating dinner. It is implausible that anything that one can do that results in these acts occurring ranks the same as any other outcome in which these acts occur. Surely, reading a very good book may be better than not reading (p. 78) any book, which in turn is better than reading a very bad book. This is something that *Value Non-Splintering* forbids.

For this reason, then, I believe the rejection of *Value Non-Splintering* is a very modest commitment about goodness. And as such, it can be shown on modest grounds that each of the principles of *Agglomeration*, *Inheritance*, and *No Conflicts* forms a fault line separating standard consequentialism (which must reject all of these claims) from other theories that can accept these claims.

3.5. Depth

How deep is this fault line? First, the fault line does not falsify *Compelling Idea* (which recall says that if the outcome of S's doing *x* is the best available outcome, then it is permissible for S to do *x*). This is because in these cases the act which has the best outcome (e.g., accepting and behaving) is obligatory and hence is permissible according to standard consequentialism and according to how cases like Zimmerman's are standardly presented.

Whether both *Left-to-Right* (i.e., if S is obligated to do *x*, then the outcome of S's doing *x* is better than the outcome of S's failing to do *x*) and *Right-to-Left* (i.e., if the outcome of S's doing *x* is better than the outcome of S's failing to do *x*, then S is obligated to do *x*) fail depends on exactly how *Value Non-Splintering* fails. To illustrate, suppose in Zimmerman's case *Accept & Behave* has an outcome that is strictly better than the outcome of *~Accept*,

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

which in turn is strictly better than the outcome of *Accept.* This is an instance of the failure of *Value Non-Splintering* that has the following structure:

> *Total Value Splintering* there is an act *A* such that $\mathbf{w_i}$ and $\mathbf{w_j}$ is an outcome in which *A* occurs and $\mathbf{w}_k$ is an outcome in which ~*A* occurs and $\mathbf{w_i}$ is strictly better than $\mathbf{w}_k$ and $\mathbf{w}_k$ is strictly better than $\mathbf{w}_j$

In cases, with this structure, both *Left-to-Right* and *Right-to-Left* fail. To see this, suppose *Inheritance Agglomeration* and *No Conflicts* hold so that it is obligatory to accept and behave, that it is obligatory to accept, and that it is not obligatory to decline. Here *Left-to-Right* fails because it is obligatory to accept even though the outcome of accepting is not strictly better than the outcome of declining. And here *Right-to-Left* fails because the outcome of declining is strictly better than the outcome of accepting but it is not obligatory to decline.

On the other hand, *Value Non-Splintering* can fail without *Total Value Splintering* holding. For example, suppose we hold that the outcome of accepting and behaving is strictly better than the outcome of declining but then only claim that the outcome of accepting (and not behaving) is equally good as the outcome of declining. This is not an instance of *Total Value* splintering because the outcome of accepting and the outcome of declining are equally good. Instead, this is an instance of the following general structure:

(p. 79)

> *Partial Value Splintering* there is an act *A* such that $\mathbf{w_i}$ and $\mathbf{w_j}$ are outcomes in which *A* occurs and $\mathbf{w}_k$ is an outcome in which ~*A* occurs and $\mathbf{w_i}$ is strictly better than $\mathbf{w}_k$ and $\mathbf{w}_k$ is equally good $\mathbf{w}_j$

As the interested reader can verify, if we only have instance of *Partial Value Splintering*, *Left-to-Right* fails, but *Right-to-Left* does not fail. Since I am skeptical there are any plausible grounds for merely accepting *Partial Value Splintering*, I will not explore this more restricted failure of *Value Non-Splintering* further. Instead, I conclude that on relatively modest grounds we have located a fault line that is a kind of intermediate between the depth of cases of deontic constraints and the shallowness of cases of deontic options.

4. Further Dimensions of Depth

There are further questions we can ask to assess the depth of a fault line. We can consider what kinds of theories are separated by a fault line and how interesting these theories are. And we can also consider what kinds of cases constitute the fault line and how interesting these cases are. We spend most of this section considering the first issue (sections 4.2–4.4), but we close with a brief discussion of the second (section 4.5).

As we will see, the cases that illustrate the failures of the principles that we are discussing separate standard consequentialism from other kinds of value-based theories. In

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

order to see this clearly, we begin by highlighting the main features of standard consequentialism.

4.1. Basic Resources

Recall that standard consequentialism is the following claim:

> S is obligated to do *x* iff the outcome of S's doing *x* is better than the outcome of S's failing to do *x*

where we say for a possible world, **w**:

> **w** is the outcome of S's doing *x* iff if S were to do *x*, then **w** would obtain

Let us look at the basic resources involved in this formulation. It, of course, involves explaining what is obligatory in terms of value. It also involves the notion of an *outcome*. As I have stipulatively defined it, an outcome is a possible world (i.e., a maximally specific way things could be), and it is the possible world that *would* result if the agent did the act.

This notion of an outcome is in one way quite broad. It is broad in the sense that it is very inclusive: the outcome of an act is not merely its causal consequences. Rather, the (p. 80) outcome of an act includes everything that would be the case. This includes things such as the act itself and events prior to the act. This broad notion of consequence is often accepted by consequentialists in order to develop the theory in the most ecumenical way possible.

But in another way, this notion of the outcome of an act involves certain strong commitments. In particular, it requires that there is a unique maximally specific way things would be if an agent performed an act. Very roughly, this amounts to a commitment to the principle of so-called conditional excluded middle (at least where the antecedent involves claims about what an agent does):

> $P \mathrel{\square\!\!\rightarrow} Q$ or $P \mathrel{\square\!\!\rightarrow} {\sim}Q$

In the standard semantics for counterfactuals, conditional excluded middle corresponds to the claim that for any *P* and any way things could be **w**, there is a unique closest possible world to **w** where *P* is true. There are ways to relax this assumption if we like, but we will adopt it for simplicity in what follows.[15]

The last feature to take note of is that this statement of consequentialism involves comparing an act's outcome with the outcome of refraining from doing the act. So standard consequentialism understands obligations in terms of the value of an outcome of an act and how it compares to the outcome of refraining from doing that act.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

4.2. Generalizing

We can locate this specific way of determining whether an act is obligatory within a more general set of theoretical options about how values determine whether an act is obligatory.

First, we can notice that consequentialism determines the deontic status of an act by considering the value of the possible world that would result if the act were performed. But we have already observed that there can be an outcome in which an act occurs that is not itself the outcome of the act. In Zimmerman's example, there is an outcome in which one accepts that is not the outcome of one accepting—the outcome in which one accepts (p. 81) the invitation and behaves well is not the outcome of accepting the invitation because one would behave poorly if one were to accept the invitation. Noticing this allows us to see that consequentialism is just one response to the question of how the deontic states of an act is related to the values of outcomes in which the act occurs. Table 4.1 summarizes a variety of positions one can take on this question, where **w** is an outcome. Consequentialism is a form of *Deontic Actualism*. It tethers the deontic status of an act to the outcome of the act. But the table shows that there are other ways in which the value of an outcome might determine whether an act is obligatory.

Table 4.1 (p. 82) How Are the Deontic Statuses of Acts Related to Outcomes in Which the Act Occurs?

Views	***O*(*A*) iff**
Deontic Maximin Possibilism	the worst **w** where *A* obtains is better than C
Deontic Maximax Possibilism	the best **w** where *A* obtains is the better than C
Deontic Actualism	the **w** that would obtain if *A* obtained is better than C

We have an open parameter C in Table 4.1. This represents different views about what the relevant comparison class is for determining whether an act ought to be done. We saw earlier what standard consequentialism looks at the outcome of not doing *A*, but again we can imagine other answers to this question such as the ones mentioned in Table 4.2.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Table 4.2 What Comparison Class Is Relevant to Determining the Deontic Status of an Act?

Views	**The comparison class for *A* is**
Better than Not	the outcome(s) relevant to ~*A*s deontic status
Better than Alt	the outcome(s) relevant to every alternative to *A*'s deontic status

Better than Alt relies on the notion of *alternative*, which we can define as follows: *B* is an alternative to *A* for an agent exactly if *A* is an act available to the agent, *B* is an act available to the agent, but the agent is not able to do *A* & *B*. It is, of course, easy to see that neither of these tables exhausts the logical space of options.[16] Rather, they simple represent a few natural ideas.

Different choices about these issues lead to different results concerning the status of our principles. Let us look at this.

4.3. Some Relations between the Choice Points and Fault Lines

Let's begin by assuming, as the standard consequentialist does, that *Deontic Actualism* is true and consider what if anything is interesting about the choice of between *Better than Not* and *Better than Alt*.

As we have seen, standard consequentialism accepts *Deontic Actualism* and *Better than Not*. And standard consequentialism is incompatible with *Inheritance*, *Agglomeration*, and *No Conflicts*. Interestingly, however, the minor variant of standard consequentialism that accepts *Deontic Actualism* but adopts *Better than Alt* is compatible with and indeed entails *Agglomeration* and *No Conflicts*. The variant of standard consequentialism that accepts *Better than Alt* does, however, shares standard consequentialism's commitment to rejecting *Inheritance*.

Though I leave an informal proof of this to Claim 4 in the appendix, we can see why we do not get a failure of these principles by returning to Zimmerman's case. There we saw that standard consequentialism entails that one is obligated to accept the invitation and behave well because this leads to the best outcome. And we saw that standard consequentialism entails that it is not the case that one is obligated to accept because the outcome of this act was worse than the outcome of not accepting. All of these claims are true according to a variant that accepts *Better than Alt* rather than *Better than Not* as well. Since the outcome of accepting and behaving is the very best one, it is better than the outcome of every alternative. Since the outcome of accepting is worse than the outcome of not accepting, the outcome of accepting is worse than some alternative. Since both

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

views claim that one is obligated to accept the invitation and behave well and that one is not obligated to accept the invitation, both reject *Inheritance*.

But standard consequentialism entails that one is obligated to decline the invitation because the outcome of declining the invitation is better than the outcome of (not not) accepting the invitation. This verdict is what leads to the failure of *Agglomeration* because the conjunction of declining the invitation and accepting the invitation and behaving is not obligatory. And this verdict also leads to the failure of *No Conflicts*.

But if we adopt *Better than Alt*, none of these results follow: the fact that the outcome of declining is better than the outcome of accepting is not sufficient to establish that one is obligated to decline. Instead, what would need to be shown is that the outcome of declining is better than the outcome of every alternative to declining. But it is easy to see that there is an alternative to declining that has a better outcome. In particular, accepting and behaving is an alternative to declining that has a better outcome. Thus, *Agglomeration* and *No Conflicts* are fault lines that separate *Deontic Actualism* paired with *Better than Not* (standard consequentialism) from *Deontic Actualism* paired with *Better than Alt*.

If we now turn to *Deontic Maximin Possibilism* when paired with *Better than Not* or *Better than Alt*, we see that this same pattern repeats itself. As Claim 5 in the appendix demonstrates, *Deontic Maximin Possibilism* paired with *Better than Not* does not validate (p. 83) *Inheritance*, *Agglomeration*, or *No Conflicts*. But, as Claim 7 in the appendix shows, *Deontic Maximin Possibilism* paired with *Better than Alt* does validate *Agglomeration* and *No Conflicts* even though it does not validate *Inheritance* (Claim 6 in the appendix).

This repeated pattern suggests the following conjecture: any (reasonable) theory that accepts *Better than Alt* validates *Agglomeration*. To prove this conjecture, we would need a more systematic grasp of the logical space of reasonable theories. Since I cannot provide such a systematic account here, I cannot prove this conjecture. But we have encountered some circumstantial evidence for it.

Corroborating this conjecture further, *Deontic Maximax Possibilism* paired with *Better than Alt* validates *Agglomeration* and *No Conflicts* (as is shown in Claim 10 in the appendix).[17] Interestingly, however, *Deontic Maximax Possibilism* paired with *Better than Alt* stands out among the theories in that it validates *Inheritance* as well (as is shown in Claim 9 in the appendix). Finally, if we turn to *Deontic Maximax Possibilism* paired with *Better than Not*, it turns out to be equivalent to *Deontic Maximax Possibilism* paired with *Better than Alt* (as is shown in Claim 8 in the appendix).

Thus, it appears *Inheritance* is a fault line separating *Deontic Maximax Possibilism* from the other theories discussed in Table 4.1 and *Agglomeration* is a fault line separating theory that accepts *Better than Alt* from other theories.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

4.4. Further Generalizations

Though we do not have the space here to explore these issues in depth, it is worth pointing out that there are still other important answers to the two questions about how deontic statuses are determined that we are exploring. If we return to the first question, there are at least following further options to consider (see Table 4.3).

Table 4.3 Further Options: How Are the Deontic Statuses of Acts Related to Outcomes in Which the Act Occurs?

Views	***O*(*A*) iff**
Deontic Maximin Fism	the worst **w** that is F and where *A* obtains is better than C
Deontic Maximax Fism	the best **w** that is F and where *A* obtains is the better than C
Deontic Averagism	some kind of average of the values of the **w**s that entail *A* is greater than C

Deontic Maximin Fism and *Deontic Maximax Fism* are like their *Possibilism* counterparts except that they place some further conditions on what the outcome where *A* obtains must be like. Some theories in the literature that fit this mold are so-called (p. 84) securantism (Portmore 2011; Ross 2012) and maximalism (Portmore 2019). These views in their *Maximax* form are typically thought to validate *Inheritance*, *Agglomeration*, and *No Conflicts*. So in this respect, they pair with their *Possibilism* counterparts. It may also be the case that rule consequentialist views (Hooker 2001) can be thought of as forms of *Deontic Maximax Fism*, though more care is required to make this assessment.

Deontic Averagism is a view that is most familiar in decision-theoretic contexts. In standard decision theory, the utility assigned to an act and whether the act ought to be done is a function of a probabilistically weighted average of the values of the various outcomes in which the act occurs. It is an interesting question for further research what properties *Deontic Averagism* may have.

Similarly, there are further options concerning comparison (see Table 4.4).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Table 4.4 Further Options: What Comparison Class Is Relevant to Determining the Deontic Status of an Act?

Views	**The comparison class for *A* is**
Better than Threshold	the outcome, **t**
Better than Context	the outcome(s) relevant to the deontic status of acts supplied by context (speaker, assessor, etc.)

Better than Threshold corresponds to simplistic forms of satisficing consequentialism (Slote 1984). And certain kinds of deontic logics that are based on preferences (Hansson 2001).

It is much harder to evaluate *Better than Context* without developing a much richer account of the role of context in determining comparison classes of acts. Luckily, there is already some important work in the literature about the semantics and logic of 'ought' and related notion that addresses some of these questions.[18]

This, then, gives us a number of avenues for future research by which we can assess and evaluate various fault lines and a variety of important theories that determine the deontic status of act by the value of outcomes.[19]

4.5. The Principles

Let us close this section by discussing the plausibility of the various principles that I have mentioned. In deontic logic, these principles are valid in the so-called Standard Deontic Logic. This is not to say that they are uncontroversial. Far from it. Indeed, the Standard (p. 85) Deontic Logic is widely rejected for a variety of reasons. I will not rehearse the challenges to Standard Deontic Logic here.[20] Instead, I will briefly illustrate the plausibility of these principles by showing how they provide tidy explanations of mundane facts about what we are obligated to do and of mundane features of our ethical thinking. This will provide some (defeasible) evidence for the principle and show how accepting or rejecting these principles is connected to broader issues.

Begin with *Inheritance*. This principle is manifest in ordinary forms of reasoning such as concluding from the fact that you ought to drive less that fifty miles per hour on a street that you ought to refrain from driving fifty-three miles per hour on that street (cf. Cariani 2013, n1). It also is closely related to forms of reasoning about what means we ought to take to our ends.

Next consider *Agglomeration*. Suppose one knows that one ought to fight in the army or perform public service and suppose one also knows that one ought to not fight in the army. In this setting, it is natural to conclude that one ought to perform public service (cf.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Horty 1993, 73). This inference is not licensed by *Inheritance* alone. But if one accepts *Agglomeration*, it follows from these two claims that one ought to both fight in the army or perform public service *and* not fight in the army. From this and *Inheritance*, it follows that one ought to perform public service. More generally, this form of reasoning is closely related to the idea that one should consider how best to achieve one's goals taken together rather than separately.

Thus, these principles, though controversial, nicely explain simple facts about obligations and simple features of ethical thought.

5. Conclusion

Let us close by briefly mentioning how the ideas that we have explored here bear on the topic of *consequentializing*, the topic of whether (and how) any nonconsequentialist theory can be given a consequentialist interpretation. There is no consensus about what follows from the fact that a nonconsequentialist theory can be given a consequentialist representation. But Jamie Dreier provides one influential answer:

> by consequentializing a theory we can keep clearer about what the important structural differences are among competing moral theories. If I am right that the consequentialist/nonconsequentialist distinction is a shallow matter of book-keeping, then consequentializing all competitors will help shine the light on distinctions that are important, like centeredness and perhaps causal versus constitutive connections between act and consequence, by clearing away the shallow differences. (2011, 115)

(p. 86) According to Dreier, important differences between ethical theories are really differences in what contributes to the value of outcomes. I believe the work we have done here casts significant doubt on Dreier's idea.[21]

To see why, consider what Dreier's idea suggests about the difference between standard consequentialism and *Deontic Maximax Possibilism*. According to Dreier, the difference that these theories have about whether to consider the outcome of *A* or the best outcome in which *A* in determining whether an act is obligatory is not a very important one. Rather, it is better to translate the *Deontic Maximax Possibilism* into a standard consequentialist format and consider what theory of value it is committed to. As we have seen, such a translation would require commitment to *Value Non-Splintering*.

Though I have no argument for this, I put it to the reader that this does not capture what is importantly different between standard consequentialism and *Deontic Maximax Possibilism*. It would be uncharitable to assume that anyone believes *Value Non-Splintering*. And I see no reason why the matter should not be taken simply at face value: these theorists can agree about the value of outcomes but disagree about how the value of various outcomes is relevant for determining the deontic status of acts.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The lesson may generalize. What many theories disagree about is not what the value of outcomes is, but the way the value of outcomes is related to the deontic status of acts. These are genuine differences that, as we have seen, can correspond to important general principles.

Thus, I conclude the project of finding and assessing fault lines in ethical theory suggests that certain strands of thought in the debate about consequentializing are mistaken. And more generally, I hope that our discussion points the way to new and interesting questions that may help us to better assess long-standing debates in moral theory.

Appendix

The proofs in this appendix are somewhat informal and rely on certain assumptions. We assume throughout that each act has a unique outcome that is a possible world, that the goodness ordering is a total ordering, that there is always a unique outcome that is better than all other outcomes, and that

$\sim O_{\perp}$: if A is inconsistent, then $\sim O(A)$ is true

Relaxing these assumptions is beyond the scope of this paper (but see n. 15 and n. 23). But for those who are interested, exploring these issues in a more general setting involves considering (p. 87) the relation between certain logical principles and certain properties of preference-like relations. There is excellent work on this topic in the deontic logic tradition.[22] But there is still room for considerable new research.

We begin by restating our key claims (sometimes in a form more amenable for the proofs to come):

Inheritance: If A entails B, then $O(A)$ entails $O(B)$

Agglomeration $O(A)$ & $O(B) \rightarrow O(A \,\&\, B)$

No Conflicts If $O(A_1)$ is true, $O(A_2)$ is true, … , and $O(A_n)$ is true, then $\{A_1, A_2, \ldots, A_n\}$ is consistent

Standard Consequentialism: $O(A)$ iff $\mathbf{w}_A$ is better than $\mathbf{w}_{\sim A}$ (where $\mathbf{w}_A$ is the outcome of doing A; similarly for other acts)

Deontic Actualism + Better than Alt: $O(A)$ iff $\mathbf{w}_A$ is better than $\mathbf{w}_B$ for each B that is an alternative to A

Deontic Maximin Possibilism+ Better than Not: $O(A)$ iff $\mathbf{w}{-}_A$ is better than $\mathbf{w}{-}_{\sim A}$ (where $\mathbf{w}{-}_A$ is the worst outcome in which A occurs; similarly for other acts)

Deontic Maximin Possibilism + Better than Alt: $O(A)$ iff $\mathbf{w}{-}_A$ is better than $\mathbf{w}{-}_B$ for each B that is an alternative to A

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Deontic Maximax Possibilism+ Better than Not: O(A) iff $\mathbf{w+}_A$ is better than $\mathbf{w+}_{\sim A}$ (where $\mathbf{w+}_A$ is the best outcome in which A occurs; similarly for other acts)

Deontic Maximax Possibilism + Better than Alt: O(A) iff $\mathbf{w+}_A$ is better than $\mathbf{w+}_B$ for each B that is an alternative to A

Value Non-Splintering if A is true at $\mathbf{w}_\mathrm{i}$ and at $\mathbf{w}_\mathbf{j}$, then for any $\mathbf{w}_\mathrm{k}$ where A is false, if $\mathbf{w}_\mathrm{i}$ is strictly better than (/worse than/equally good as) $\mathbf{w}_\mathrm{k}$, then $\mathbf{w}_\mathrm{j}$ is better than (/worse than/equally good as) $\mathbf{w}_\mathrm{k}$ (where $\mathbf{w}_\mathrm{i}$, $\mathbf{w}_\mathrm{j}$, $\mathbf{w}_\mathrm{k}$ are outcomes)

We now prove each of the main claims made in the text.

Claim 1 *Standard Consequentialism* and *Value Non-Splintering* entail *Inheritance*

Assume *Standard Consequentialism and Value Non-Splittering* are true and suppose for *reductio* that A entails B, O(A), but ~O(B). Since O(A), *Standard Consequentialism* entails $\mathbf{w}_A$ is better than $\mathbf{w}_{\sim A}$. Since A entails B, B is true at $\mathbf{w}_A$. By *Value-Non-Splintering*,$\mathbf{w}_B$ is better than $\mathbf{w}_{\sim A}$.

Since ~O(B), *Standard Consequentialism* entails $\mathbf{w}_B$ is not better than $\mathbf{w}_{\sim B}$ (i.e., $\mathbf{w}_B$ is equally good or worse than $\mathbf{w}_{\sim B}$). Since A entails B, ~A is true at $\mathbf{w}_{\sim B}$. By *Value Non-Splintering*, $\mathbf{w}_B$ is not better than $\mathbf{w}_{\sim A}$ (i.e., $\mathbf{w}_B$ is equally good or worse than $\mathbf{w}_{\sim A}$). Contradiction.

Thus *Standard Consequentialism* and *Value Non-Splintering* entail *Inheritance*

Claim 2 *Standard Consequentialism* and *Value Non-Splintering* entail *Agglomeration*

Assume *Standard Consequentialism and Value Non-Splittering* are true and suppose for *reductio* O(A),O(B), but ~O(A & B). By *Standard Consequentialism*, $\mathbf{w}_A$ is better than $\mathbf{w}_{\sim A}$ and $\mathbf{w}_B$ is better than $\mathbf{w}_{\sim B}$, Since A and B are both true at $\mathbf{w}_{(A\ \&\ B)}$, *Value Non-Splintering* entails that $\mathbf{w}_{(A\ \&\ B)}$ is better than $\mathbf{w}_{\sim A}$ and $\mathbf{w}_{(A\ \&\ B)}$ is better than $\mathbf{w}_{\sim B}$

(p. 88) Since ~O(A & B), *Standard Consequentialism* entails that $\mathbf{w}_{(A\ \&\ B)}$ is not better than $\mathbf{w}_{\sim(A\ \&\ B)}$ (i.e., $\mathbf{w}_{(A\ \&\ B)}$ is either equally good or worse than $\mathbf{w}_{\sim(A\ \&\ B)}$). Now either (i) ~$A$ is true at $\mathbf{w}_{\sim(A\ \&\ B)}$ or (ii) ~B is true at $\mathbf{w}_{\sim(A\ \&\ B)}$. Suppose (i). Then by *Value Non-Splintering*$\mathbf{w}_{(A\ \&\ B)}$ is not better than $\mathbf{w}_{\sim A}$. Contradiction. Suppose instead (ii). Then by *Value Non-Splintering*$\mathbf{w}_{(A\ \&\ B)}$ is not better than $\mathbf{w}_{\sim B}$. Contradiction.

Thus, *Standard Consequentialism* and *Value Non-Splintering* entail *Agglomeration.*

Claim 3 *Standard Consequentialism* and *Value Non-Splintering* entail *No Conflicts*

Assume *Standard Consequentialism* and *Value Non-Splintering* are true and suppose for *reductio* that O(A_1), O(A_2), … , O(A_n) are true but {A_1, A_2, … , A_n} is incon-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

sistent. By **Claim 2** and O(A_1), O(A_2), … , O(A_n), O(A_1&A_2& … & A_n). But given ~O$_\perp$ and the assumption that{A_1, A_2, … , A_n} is inconsistent, ~O(A_1&A_2& … & A_n). Contradiction.

Thus, *Standard Consequentialism* and *Value Non-Splintering* entail *No Conflicts*.

Claim 4 *Deontic Actualism + Better than Alt* entails *Agglomeration*

Assume *Deontic Actualism + Better than Alt* is true and suppose for *reductio* O(A) and O(B) but ~O(A & B). Given *Deontic Actualism + Better than Alt*, there is some act C that is an alternative to A & B such that $\mathbf{w}_{A\ \&\ B}$ is not better than $\mathbf{w}_C$. Since C is an alternative to A & B, it follows either that (i) ~A & C is true at $\mathbf{w}_C$ or that (ii) ~B & C is true at $\mathbf{w}_C$.

Suppose (i). It follows that $\mathbf{w}_{(\sim A\ \&\ C)} = \mathbf{w}_C$.[23] Since O($A$) and ~$A$ & C is an alternative to A, *Deontic Actualism + Better than Alt* entail $\mathbf{w}_A$ is better than $\mathbf{w}_{(\sim A\ \&\ C)} = \mathbf{w}_C$. Now either A & B is true at $\mathbf{w}_A$ or it is not.

Suppose A & B is true at $\mathbf{w}_A$. It follows that $\mathbf{w}_A = \mathrm{w}_{(A\ \&\ B)}$ so $\mathbf{w}_{(A\ \&\ B)}$ is better than $\mathbf{w}_C$. This contradicts our earlier claim that $\mathrm{w}_{(A\ \&\ B)}$ is not better than $\mathbf{w}_C$.

Suppose instead A & B is not true at $\mathbf{w}_A$. It follows $\mathrm{w}_{(A\ \&\ \sim B)} = \mathbf{w}_A$. Since A & ~B is an alternative to B and O(B), *Deontic Actualism + Better than Alt* entails $\mathbf{w}_B$ is better than $\mathbf{w}_{(A\ \&\ \sim B)} = \mathbf{w}_A$. Either A & B is true at $\mathbf{w}_B$ or it isn't. If it is, then $\mathbf{w}_B = \mathbf{w}_{(A\ \&\ B)}$ and so $\mathbf{w}_{(A\ \&\ B)}$ is better than $\mathbf{w}_A$ which is better than $\mathbf{w}_C$. This contradicts the assumption that $\mathbf{w}_{(A\ \&\ B)}$ is not better than $\mathbf{w}_C$. So it must be that A & B is false at $\mathbf{w}_B$. So $\mathbf{w}_B = \mathbf{w}_{(\sim A\ \&\ B)}$. Since ~$A$ & B is an alternative to A and O(A), *Deontic Actualism + Better than Alt*, $\mathbf{w}_A$ is better than $\mathbf{w}_{(\sim A\ \&\ B)} = \mathbf{w}_B$ which contradicts our assumption that $\mathbf{w}_B$ is better than $\mathbf{w}_{(A\ \&\ \sim B)} = \mathbf{w}_A$. Thus, (i) cannot hold.

Suppose instead, then, that (ii) holds. Analogous reasoning shows that (ii) cannot be true.

Thus, *Deontic Actualism + Better than Alt* entails *Agglomeration*. As a corollary of this, *Deontic Actualism + Better than Alt* also entails *No Conflicts* for analogous reasons to those given for **Claim 3**.

(p. 89) **Claim 5** *Deontic Maximin Possibilism + Better than Not* does not validate *Inheritance*, *Agglomeration*, or *No Conflicts*

Consider a four-world model where the numbers are understood to represent the value of each world:

$\mathbf{w}_{A\ \&\ B}$: 100 $\mathbf{w}_{A\ \&\ \sim B}$: 25 $\mathbf{w}_{\sim A\ \&\ B}$: 50 $\mathbf{w}_{\sim A\ \&\ \sim B}$: 50

Since $\mathbf{w}\text{-}_{(A\ \&\ B)}$ is better than $\mathbf{w}\text{-}_{\sim(A\ \&\ B)}$, *Deontic Maximin Possibilism + Better than Not* entails O(A & B). But $\mathbf{w}\text{-}_A$ is not better than $\mathbf{w}\text{-}_{\sim A}$, so *Deontic Maximin Possibilism + Better than Not* entails ~O(A). Thus, *Inheritance* does not hold.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Furthermore, since $\mathbf{w}\text{-}_{\sim A}$ is better than $\mathbf{w}\text{-}_{A}$, *Deontic Maximin Possibilism + Better than Not* entails O(∼*A*). We have already seen that O(*A* & *B*). Since {*A* & *B*, ∼*A*} is inconsistent, *No Conflicts* does not hold.

Finally, we are assuming ∼O(*f*), where *f* is any inconsistent claim, so ∼O(*A* & *B* & ∼*A*). Thus, *Agglomeration* does not hold.

Claim 6 *DeonticMaximin Possibilism + Better than Alt* does not validate *Inheritance*

Using the same model as **Claim 5**, $\mathbf{w}\text{-}_{(A\ \&\ B)}$ is better than all worlds, so $\mathbf{w}\text{-}_{(A\ \&\ B)}$ is better than $\mathbf{w}\text{-}_{C}$ for any *C* that is an alternative to *A* & *B*. By *Deontic Maximin Possibilism + Better than Alt*, O(*A* & *B*). But once again $\mathbf{w}\text{-}_{A}$ is not better than $\mathbf{w}\text{-}_{\sim A}$ and ∼*A* is an alternative to *A*, so *Deontic Maximin Possibilism + Better than Alt* entails ∼O(*A*). Thus, *Inheritance* does not hold.

Claim 7 *Deontic Maximin Possibilism + Better than Alt* entails *Agglomeration*

We begin with a lemma.

Lemma 7.1 *Deontic Maximin Possibilism + Better than Alt* entails that if O(*A*) and O(*B*), then either *A* entails *B* or *B* entails *A*.

Assume *Deontic Actualism + Better than Alt* is true and suppose for *reductio* O(*A*), O(*B*), but *A* does not entail *B* and *B* does not entail *A*. Given that *A* does not entail *B* and *B* does not entail *A*, it follows that there is a $\mathbf{w}\text{-}_{(\sim A\ \&\ B)}$ and there is a $\mathbf{w}\text{-}_{(A\ \&\ \sim B)}$.

Given O(A), *Deontic Maximin Possibilism + Better than Alt,* and the fact that ∼*A* & *B* is an alternative to *A*, it follows that $\mathbf{w}\text{-}_{A}$ is better than $\mathbf{w}\text{-}_{(\sim A\ \&\ B)}$. And similarly, $\mathrm{w}\text{-}_{B}$ is better than $\mathbf{w}\text{-}_{(A\ \&\ \sim B)}$.

But if $\mathbf{w}\text{-}_{A}$ is better than $\mathbf{w}\text{-}_{(\sim A\ \&\ B)}$, then $\mathbf{w}\text{-}_{(A\ \&\ \sim B)}$ is better than $\mathbf{w}\text{-}_{(\sim A\ \&\ B)}$. And similarly, $\mathrm{w}\text{-}_{(\sim A\ \&\ B)}$ is better than $\mathbf{w}\text{-}_{(A\ \&\ \sim B)}$. This is a contradiction, so the lemma holds.

We now turn to the main proof. Assume *Deontic Actualism + Better than Alt* is true and assume for *reductio* O(*A*), O(*B*), but ∼O(*A* & *B*). Given *Deontic Maximin Possibilism + Better than Alt*, there is an alternative *C* to *A* & *B* such that $\mathbf{w}\text{-}_{(A\ \&\ B)}$ is not better than $\mathbf{w}\text{-}_{C}$.

By Lemma 7.1 either *A* entails *B* or *B* entails *A*. Suppose *A* entails *B* so *A* and *A* & *B* are equivalents and so every alternative to *A* is an alternative to *A* & *B* and vice-versa and $\mathbf{w}\text{-}_{A} = \mathbf{w}\text{-}_{(A\ \&\ B)}$. Since O(*A*), *Deontic Maximin Possibilism + Better than Alt* entails that $\mathbf{w}\text{-}_{A}$ is better than $\mathbf{w}\text{-}_{C}$ for any *C* that is an alternative to *A*. Thus, $\mathbf{w}\text{-}_{(A\ \&\ B)}$ is better than $\mathbf{w}\text{-}_{C}$ for any *C* that is an alternative to *A* & *B*. This is a contradiction.

So suppose *B* entails *A*. Analogous reasoning shows we reach a contradiction.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Thus, *Deontic Maximin Possibilism + Better than Alt* entails *Agglomeration*.

As a corollary of this, *Deontic Maximin Possibilism + Better than Alt* also entails *No Conflicts* for analogous reasons to those given for **Claim 3**.

Claim 8 *Deontic Maximax Possibilism + Better than Alt* and *Deontic Maximax Possibilism + Better than Not* give equivalent verdicts about what is obligatory

This claim trivially follows from the following two lemmas.

(p. 90) Lemma 8.1 If O(*A*) is true according to *Deontic Maximax Possibilism + Better than Alt* entails that O(*A*), then O(*A*) is true according *Deontic Maximax Possibilism + Better than Not*

Suppose O(*A*) is true according to *Deontic Maximax Possibilism + Better than Alt*. Thus, $\mathbf{w+}_A$ is better than $\mathbf{w+}_C$ for every alternative *C* to *A*. ~*A* is an alternative to *A*. Thus, $\mathbf{w+}_A$ is better than $\mathbf{w+}_{\sim A}$ is true. So according to *Deontic Maximax Possibilism + Better than Not*, O(*A*) is true.

Lemma 8.2 If O(*A*) is true according to *Deontic Maximax Possibilism + Better than Not*, then O(*A*) is true according to *Deontic Maximax Possibilism + Better than Alt*

Suppose O(*A*) is true according to *Deontic Maximax Possibilism + Better than Not*. Thus, $\mathbf{w+}_A$ is better than $\mathbf{w+}_{\sim A}$. Now for any $\mathbf{w+}_C$ such that *C* is an alternative to *A*, ~*A* is true at $\mathbf{w+}_C$. So $\mathbf{w+}_{\sim A}$ is at least as good as $\mathbf{w+}_C$. Thus, for any alternative *C* to *A*, $\mathbf{w+}_A$ is better than $\mathbf{w+}_C$. So according to *Deontic Maximax Possibilism + Better than Alt*, O(*A*) is true.

Claim 9 *Deontic Maximax Possibilism + Better than Alt* entails *Inheritance*

Suppose *Deontic Maximax Possibilism + Better than Alt* is true and assume *A* entails *B* and O(*A*). By *Deontic Maximax Possibilism + Better than Alt*, $\mathbf{w+}_A$ is better than $\mathbf{w+}_C$ for any alternative *C* to *A*. Since *A* entails *B*, *B* is true at $\mathbf{w+}_A$. Therefore, $\mathbf{w+}_B$ is at least as good as $\mathbf{w+}_A$. Next if *C'* is an alternative to *B*, *C'* is an alternative to *A* since *A* entails *B*. Thus, $\mathbf{w+}_A$ is better than $\mathbf{w+}_{C'}$ for any alternative *C'* to *B*. So since $\mathbf{w+}_B$ is at least as good as $\mathbf{w+}_A$, $\mathbf{w+}_B$ is better than $\mathbf{w+}_{C'}$ for any alternative *C'* to *B*. So by *Deontic Maximax Possibilism + Better than Alt*, O(*B*). Therefore, *InheritanceDeontic Maximax Possibilism + Better than Alt* entails *Inheritance*

Claim 10 *Deontic Maximax Possibilism + Better than Alt* entails *Agglomeration*

We begin with a lemma.

Lemma 10.1 *Deontic Maximax Possibilism + Better than Alt* entails that if O(*A*) and O(*B*), then $\mathbf{w+}_A = \mathbf{w+}_{(A\ \&\ B)} = \mathbf{w+}_B$.

Assume Deontic Maximax Possibilism + Better than Alt is true and suppose O(*A*) and O(*B*). Given *Deontic Maximax Possibilism + Better than Alt* and O(*A*), $\mathbf{w+}_A$ is

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

better than $\mathbf{w+}_C$ for any alternative C to A. Either A & B is true at $\mathbf{w+}_A$ or it isn't. If A & B is true at $\mathbf{w+}_A$, then $\mathbf{w+}_A = \mathbf{w+}_{(A \& B)}$.

Suppose instead A & B is false at $\mathbf{w+}_A$. Thus, A & $\sim B$ is true at $\mathbf{w+}_A$ and $\mathbf{w+}_A = \mathbf{w+}_{(A \& \sim B)}$. Since A & $\sim B$ is an alternative to B and O(B), *Deontic Maximax Possibilism + Better than Alt* entail that $\mathbf{w+}_B$ is better than $\mathbf{w+}_{(A \& \sim B)} = \mathbf{w+}_A$. Now either A is true at $\mathbf{w+}_B$ or it isn't.

Suppose A is true at $\mathbf{w+}_B$. It follows that since $\mathbf{w+}_B$ is better than $\mathbf{w+}_{(A \& \sim B))}$, $\mathbf{w+}_{(A \& \sim B)} \neq \mathbf{w+}_A$. This contradicts our previous claim that $\mathbf{w+}_A = \mathbf{w+}_{(A \& \sim B)}$.

Suppose then A is false and so $\mathbf{w+}_{(\sim A \& B)} = \mathbf{w+}_B$. Since $\sim A$ & B is an alternative to A and since $\mathbf{w+}_{(\sim A \& B)} = \mathbf{w+}_B$ is better than $\mathbf{w+}_A$. So *Deontic Maximax Possibilism + Better than Alt* entails $\sim$O(A) which contradicts our assumption that O(A). So A & B must be true at $\mathbf{w+}_A$. So $\mathbf{w+}_{(A \& B)} = \mathbf{w+}_A$.

Next, given *Deontic Maximax Possibilism + Better than Alt* and O(B), $\mathbf{w+}_B$ is better than $\mathbf{w+}_C$ for any alternative C to B. Either A & B is true at $\mathbf{w+}_B$ or it isn't. By analogous reasoning, we establish A & B is true at $\mathbf{w+}_B$ and therefore, $\mathbf{w+}_{(A \& B)} = \mathbf{w+}_B$.

Thus, $\mathbf{w+}_A = \mathbf{w+}_{(A \& B)} = \mathbf{w+}_B$ so Lemma 10.1 holds.

We now turn to the main proof. Assume *Deontic Maximax Possibilism + Better than Alt* is true and suppose for *reductio* O(A) and O(B), but $\sim$O(A & B). By *Deontic Maximax Possibilism + Better than Alt*, there is C that is an alternative to A & B such that $\mathbf{w+}_{(A \& B)}$ is (p. 91) not better than $\mathbf{w+}_C$. Since C is an alternative to A & B, either (i) $\sim A$& C is true at $\mathbf{w+}_C$ or (ii) $\sim B$ & C is true at $\mathbf{w+}_C$.

Suppose (i). So $\mathbf{w+}_C = \mathbf{w+}_{(\sim A \& C)}$. Given O($A$) and the fact that $\sim A$ & C is an alternative to A, *Deontic Maximax Possibilism + Better than Alt* entails that $\mathbf{w+}_A$ is better than $\mathbf{w+}_{(\sim A \& C)} = \mathbf{w+}_C$. But by Lemma 10.1, $\mathbf{w+}_A = \mathbf{w+}_{(A \& B)}$ so $\mathbf{w+}_{(A \& B)}$ is better than $\mathbf{w+}_{(\sim A \& C)} = \mathbf{w+}_C$. This contradicts our earlier claim that $\mathbf{w+}_{(A \& B)}$ is not better than $\mathbf{w+}_C$. So (i) must be false.

Suppose (ii). Analogous reasoning establishes that (ii) must be false.

Thus, *Deontic Maximax Possibilism + Better than Alt* entails *Agglomeration.*

As a corollary of this, *Deontic Maximax Possibilism + Better than Alt* also entails *No Conflicts* for analogous reasons to those given for **Claim 3**.

References

Brown, Campbell. 2011. "Consequentialize This." *Ethics* 121, no. 4: 749–771.

Brown, Campbell. 2018. "Maximalism and the Structure of Acts." *Nous* 52, no. 4: 752–771.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Cariani, Fabrizio. 2013. “*Ought* and Resolution Semantics.” *Nous* 47, no. 3: 534–558.

Cariani, Fabrizio. 2016. “Consequence and Contrast in Deontic Semantics.” *Journal of Philosophy* 113, no. 8: 396–416.

Darwall, Stephen. 1986. “Agent-Centered Restrictions from the Inside Out.” *Philosophical Studies* 50, no. 3: 291–319.

Dietrich, Franz, and List, Christian. 2017. “What Matters and How It Matters.” *Philosophical Review* 126, no. 4: 421–479.

Dreier, James. 1993. “Structures of Normative Theories.” *Monist* 76, no. 1: 22–40.

Dreier, James. 2011. “In Defense of Consequentializing.” In *Oxford Studies in Normative Ethics 1*, edited by Mark Timmons, 97–119. Oxford: Oxford University Press.

Foot, Phillipa. 1978. “The Problem of Abortion and the Doctrine of Double Effect.” In *Virtues and Vices*, 19–32. Berkeley: University of California Press.

Goble, Lou. 1990a. “A Logic of *good*, *should*, and *would:* Part I.” *Journal of Philosophical Logic* 19, no. 2: 169–199.

Goble, Lou. 1990b. “A Logic of *good*, *should*, and *would:* Part II.” *Journal of Philosophical Logic* 19, no. 3: 253–276.

Hansson, Sven Ove. 2001. *The Structure of Values and Norms*. New York: Cambridge University Press.

Hilpinen, Risto, and McNamara, Paul. 2014. “Deontic Logic.” In *Handbook of Deontic Logic and Normative Systems 1*, edited by Dov Gabbay, John Horty, Xavier Parent, Ron van der Meyden, and Leon van der Torre, 3–136. Milton Keynes, OH: College Publications.

Hooker, Brad. 2001. *Ideal Code, Real World*. Oxford: Oxford University Press.

Horty, John. 1993. “Deontic Logic as Founded on Nonmonotonic Logic.” *Annals of Mathematics and Artificial Intelligence* 9, no. 1–2: 69–91.

Jackson, Frank, and Pargetter, Robert. 1986. “Oughts, Options, and Actualism.” *The Philosophical Review* 95, no. 2: 233–255.

Lewis, David. (1973). *Counterfactuals*. Cambridge: Harvard University Press.

Lousie, Jennie. 2004. “Relativity of Value and the Consequentialist Umbrella.” *The Philosophical Quarterly* 54, no. 217: 518–536.

McNamara, Paul. 2019. “Deontic Logic.” In *The Stanford Encyclopedia of Philosophy* (Summer 2019 edition), edited by Edward Zalta. **https://plato.stanford.edu/archives/sum2019/entries/logic-deontic/**.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(p. 92) Nair, Shyam. 2014. “A Fault Line in Ethical Theory.” *Philosophical Perspectives* 28, no. 1: 173–200.

Oddie, Graham, and Milne, Peter. 1991. “Act and Value.” *Theoria* 57, no. 1–2: 42–76.

Pearl, Judea. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Portmore, Douglas. 2007. “Consequentializing Moral Theories.” *Pacific Philosophical Quarterly* 88, no. 1: 39–73.

Portmore, Douglas. 2009. “Consequentializing.” *Philosophy Compass* 4, no. 2: 329–347.

Portmore, Douglas. 2011. *Commonsense Consequentialism*. New York: Oxford University Press.

Portmore, Douglas. 2019. *Opting for Best*. New York: Oxford University Press.

Ross, Jacob. 2012. “Actualism, Possibilism, and Beyond.” In *Oxford Studies in Normative Ethics 2*, edited by Mark Timmons, 74–96. Oxford: Oxford University Press.

Scheffler, Samuel. 2003. *The Rejection of Consequentialism.* Revised ed. Oxford: Oxford University Press.

Schroeder, Mark. 2007. “Teleology, Agent-Relative Value, and “Good’.” *Ethics* 117, no. 2: 265–295.

Slote, Michael. 1984. “Satisficing Consequentialism.” *Proceedings of Aristotelian Society* 58, no. 1: 139–163.

Snedegar, Justin. 2017. *Contrastive Reasons*. Oxford: Oxford University Press.

Wedgwood, Ralph. 2009. “Intrinsic Values and Reasons for Action.” *Philosophical Issues* 19, no. 1: 342–363.

Williams, Bernard. 1973. “A Critique of Utilitarianism.” In *Utilitarianism: For and Against*, edited by J. J. C. Smart and Bernard Williams, 75–150. Cambridge: Cambridge University Press.

Zimmerman, Michael. 2006. *The Concept of Moral Obligation*. Cambridge: Cambridge University Press.

Notes:

([Thanks to Doug Portmore for comments on this article.])

([2]) Important proponents of the approach include Dreier (1993, 2011); Lousie (2004); Oddie and Milne (1991); and Portmore (2007, 2009). Important criticisms include Brown

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(2011) and Schroeder (2007). An important recent work that is relevant to this discussion but harder to categorize is Dietrich and List (2017).

([3]) This section is heavily informed by Nair (2014). The remaining sections discuss new material.

([4]) These cases are sometimes called cases of agent-centered restrictions and are taken to show that consequentialism only allows agent-neutral considerations to be relevant. The classic discussion in this vein is perhaps Scheffler (2003) (see also Darwall [1986], which, among other things, provides a useful historical and conceptual background for understanding these cases).

([5]) I stipulate that S's refraining from doing *x* is S not doing *x*.

([6]) This neutral argument is essentially the one in Brown (2011, 761–763).

([7]) Though it abuses language, I treat 'ought' and 'obligatory' and their variants as synonymous throughout this paper.

([8]) Roughly, the argument proceeds by claiming that cases of deontic constraints only arise when the preferences that agents ought to have differ. A subclass of such cases will be cases with the structure of a prisoner's dilemma. In this subclass, the *Unanimity* principle forces the "cooperate-cooperate" outcome to be ranked ahead of the "defect-defect" outcome in such a way that allows us to show that consequentialism is deeply incompatible with cases of deontic constraints of this structure.

([9]) I stipulate "S refrains from doing *x*" is equivalent to the sentential negation of "S does *x*."

([10]) Cases of this sort are discussed in the so-called actualism/possibilism debate. See Cohen and Timmerman, Chapter 7, this volume (and the citations therein) for a rich discussion of this literature. There are many interesting relations between their discussion and what is to follow. Unfortunately, I am unable to pursue these at this time.

([11]) This can be justified by the Kantian idea that one is obligated to do something only if one can do it or by the assumption that an impossible situation is worse than any possible situation.

([12]) *No Strict Conflicts* is something that standard consequentialism accepts. This is because *A* cannot have an outcome that is better than the outcome of ~*A* while ~*A* has a better outcome than the outcome of *A* at the same time.

([13]) Others such as Brown (2018) and Portmore (2019, chap. 4) have also thought of this case (or cases similar to it) as a case where there are multiple versions or ways of doing an act.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([14]) The qualification that "$\mathbf{w}_i$, $\mathbf{w}_j$, $\mathbf{w}_k$ are outcomes" restricts our attention to possible worlds that would result from some act available to the agent. So the principle does not apply to possible worlds that could not result from some act available to the agent.

([15]) Here is a more general definition of an outcome where an outcome, *o,* is a proposition (a set of possible worlds):

> *o* is the outcome of S's doing *x* iff *o* is true at exactly those **w** such that if S were to do *x*, then **w** might obtain

If this more general characterization is in place, we face further questions. We must ask how the value of *o* is related to the value of each **w**. We do not explore this issue here (though some of our discussion below indirectly bears on this issue). Though I cannot pursue this here, the main claims in the appendix can be shown to hold in this more general setting if we make an additional assumption. If o_1, o_2, … , o_n are cells of a partition of the set *o*, then the assumption we need is that it is not the case that *o* is strictly better than o_i for each *i* such that $1 \leq i \leq n$ and it is not the case that *o* is strictly worse than o_i for each *i* such that $1 \leq i \leq n$.

([16]) For example, Table 4.2 is parasitic on what choice a theory makes in Table 4.1 so that there is a kind of matching evaluation between *A* and its comparison class. But another logically possible approach is to use unmatched comparisons. For example, there is a theory that accepts *Deontic Maximax Possibilism* with a comparison class to the outcome of ~*A* (rather than compared to the best outcome in which ~*A*). Though worthy of further exploration, I do not consider these unmatched approaches here.

([17]) The interested reader can also check that the proof strategy for each of these results is remarkably similar: each proof exploits the fact that ~*A* & *B* and *A* & ~*B* are alternatives to *A* and *B*, respectively, to establish that O(*A* & *B)* given O(*A*) and O(*B*).

([18]) The role of context arises in discussion of cases like Zimmerman's in Jackson and Pargetter (1986). But see Snedegar (2017) and Cariani (2013, 2016) for contemporary discussion.

([19]) There is a further question as to the relationship between the value of an act and the value of outcome. Most theorists accept *Value Actualism* according to which the value of an act is determined by the value of its outcome. But some (e.g., Wedgwood 2009) do not accept this view.

([20]) See McNamara (2019) and especially Hilpinen and McNamara (2014) for synoptic discussions of Standard Deontic Logic, challenges to various principles including the ones discussed here, and a broader sense of the state of play in deontic logic.

([21]) It may be that Dreier's remarks are not intended to apply to the context discussed in the next paragraph. If so, my remarks are not a direct criticism of his view. That said, I believe that discussion here provides materials for objecting to the use of his ideas in the

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

context he clearly intends. See also Hurley, Chapter 2, this volume, for a distinct criticism of Dreier's idea.

([22]) See especially Goble (1990a, 1990b) and Hansson (2001).

([23]) The reasoning that justifies this inference can be made more explicit. We begin by inferring 'C & $\sim A$ & $C \boxrightarrow \mathbf{w}_C$' from '$C \boxrightarrow \sim A$ & C' and '$C \boxrightarrow \mathbf{w}_C$'. This is an instance of the so-called cautious monotonicity or composition principle about counterfactuals that says that if $P \boxrightarrow Q$ and $P \boxrightarrow R$, then P& $Q \boxrightarrow R$. Next we infer '$\sim A$ & $C \boxrightarrow \mathbf{w}_C$' from '$C$ & $\sim A$ & $C \boxrightarrow \mathbf{w}_C$'. This is an instance of the principle that if P and Q are logically equivalent, then $P \boxrightarrow R$ iff $Q \boxrightarrow R$. Both of these principles hold in popular accounts of counterfactuals such as those due to Lewis 1973 and due to Pearl 2009 (and they do not depend on the assumption that the outcome of each act is a unique possible world).

Finally, by definition we know that $\sim A$ & $C \boxrightarrow \mathbf{w}_{(\sim A \,\&\, C)}$. Since $\mathbf{w}_{(\sim A \,\&\, C)}$ and $\mathbf{w}_C$ are maximally specific ways the world could be, it follows therefore that $\mathbf{w}_{(\sim A \,\&\, C)} = \mathbf{w}_C$. Though I suppress these details from now on, this form of argument is relied on in several places in this appendix. (The notion of outcome stated in n. 15 allows that an outcome need not be a single possible world. Though I will not present the proof here, the last step in this reasoning can be reconstructed in that more general setting as well.)

Shyam Nair

Shyam Nair is Assistant Professor of Philosophy at Arizona State University. He is primarily interested in issues in ethics, epistemology, and philosophical logic. His research focuses on formal and philosophical questions at the intersection of these fields concerning how best to model what we ought to do, what we ought to believe, and how we ought to reason.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The Alienation Objection to Consequentialism

Calvin C. Baker and Barry Maguire

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.25

Abstract and Keywords

An ethical theory is alienating if accepting the theory inhibits the agent from fitting participation in some normative ideal, such as some ideal of integrity, friendship, or community. Many normative ideals involve nonconsequentialist behavior of some form or another. If such ideals are normatively authoritative, they constitute counterexamples to consequentialism unless their authority can be explained or explained away. We address a range of attempts to avoid such counterexamples and argue that consequentialism cannot by itself account for the normative authority of all plausible such ideals. At best, consequentialism can find a more modest place in an ethical theory that includes nonconsequentialist principles with their own normative authority.

Keywords: alienation, commitment, motives, normative ideals, authority

1. Alienation

THE state of alienation is rather complex.[1],[2] An agential *subject* is alienated from some *object*—perhaps a friend, a stranger, a group, or her own self[3]—when she is inhibited from fitting participation in some pertinent normative ideal involving the object. The normative ideal will often involve a relationship of some kind, for instance, being a friend, a teacher, or a fellow citizen; there are also ideals concerning the relations between one's own ethical principles, commitments, affects, and motives. You might be alienated from your friend by your acceptance of overly formal norms concerning friendly intercourse; or from your community by the prevalence of an ideology of individual achievement; or from yourself by your failure to be motivated by your own averred ethical ideals.

Not everything plausibly thought of as alienation involves this structure (cf. Schacht 1970 and Leopold 2018). Another important theme in historical discussions concerns the absence of control: whether over political institutions (e.g., in Hobbes or Rousseau) or civil society (e.g., in Hegel), or the patterns and outputs of economic and social production (p. 402) (e.g., in Marx), or in the aesthetic quality of our lives (e.g., in Schiller and Marcuse). However, we do think there is a core idea, strongly associated with the notion of

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

alienation, involving the inhibition of participation in some normative ideal for a kind of relationship. This will be our focus.

We will restrict our attention to normative ideals that involve some sort of *harmony, connection, or closeness* between persons or groups. Such relationships involve standards of fittingness bearing on one's participation (cf. Wood 1999 and Leopold 2018). In this sense, the restricted concept of a normative ideal that concerns us involves its own distinctive normative standards. But these standards are not, as such, *authoritatively* normative, in the sense of necessarily bearing on what one ought to do, all things considered.[4] This is clear from consideration of harmful ideals: consider the just alienated from the racist society; or the socialist alienated in the libertarian society. In such cases, there may be no authoritative reason for the alienated agent to fit into the relevant normative ideal; rather there is authoritative reason for the prevailing norms to change.[5]

So, it does not quite follow that alienation is, as such, undesirable. For one thing, undesirability is an axiological property and nonauthoritativeness is a deontic property. It is so far an open question whether there are normative ideals the realization of which would be good but the standards of which are not authoritative, for instance if their normativity is defeated. This is not too far-fetched: realizing the second-best normative ideals might bring about a great deal of value, but nevertheless it may be that the availability of a better ideal, other things being equal, renders the better standards authoritatively normative instead.

The more pertinent question is the following: Are there any normative ideals that are authoritative, the authority of which does not depend upon consequentialist assessment? To sharpen the question a little, say that for an ethical principle to be ethically *fundamental* is for it to be authoritative and not ethically explained by anything else (cf. Maguire 2015). For an ethical principle to be *supreme* is for it to be the only ethically fundamental principle. If there is at least one authoritative normative ideal whose authority does not depend upon whether its realization meets the test of consequentialist assessment, then Supreme Consequentialism—the thesis that the consequentialist standard of assessment is the only fundamental ethical principle—is false. It seems extremely plausible that there is a range of normative ideals—concerning varieties of friendship, community, and integrity—that are authoritative in this way. If so, it is a cost to a consequentialist theory if it is unable to instruct agents to properly accord with them.

(p. 403) How much of a cost? We lack space to address this methodological question. We will just say this. It is plausible that some authoritative normative ideals constitute inputs to ethical theorizing that any satisfactory ethical theory needs to accommodate. Supreme Consequentialism does not have this default epistemic status. It purports to *be* the final ethical theory—the output of this process of ethical theorizing. Consequently, if Supreme Consequentialism conflicts with such normative ideals, it is Supreme Consequentialism that should be revised, rather than the normative ideals.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Our question, then, is whether this or that "consequentialist" ethical theory yields the result that an agent ought to properly accord with this or that normative ideal. We will consider four classes of ethical theories: Hybrid Theories, Relative Value Theories, Global Consequentialism, and Leveled Consequentialism.[6] We will argue that there are some plausibly authoritative normative ideals that cannot be accommodated by the consequentialist principle in any of them. It follows that any version of Supreme Consequentialism is alienating relative to these ideals.

2. The Alienation Challenge Facing Consequentialist Theories

By *consequentialism* we mean an ethical principle according to which pertinent deontic facts about something are fully explained by facts about the value promoted by that thing and its alternatives, or by some set of suitably related alternatives. A version of consequentialism is *direct* if deontic facts about alternatives are explained by the values promoted by those alternatives, and *indirect* if explained by the values promoted by some suitably related alternatives.[7]

Let Harmonious Act Consequentialism be the conjunction of direct Act Consequentialism, according to which one ought to perform an action if and only if and because it promotes more value than any alternative, and the principle that one ought to be motivated, in taking an option, by whatever explains why one ought to do so (cf. Arpaly 2011 and Way 2017).

The simplest alienation objection runs as follows. Suppose that taking some option that benefits Joe is the alternative that promotes the most value. You take the option, and you are motivated by the fact that doing so promotes more value than any alternative. Suppose that Joe is a close friend of yours. According to a plausible normative ideal of (p. 404) friendship, it is fitting to be motivated to help friends by the fact that one's action will help one's friend, not by the fact that it will promote the most value overall.[8] The objection goes that one's acceptance[9] of Harmonious Act Consequentialism, insofar as it ensures that one lacks the former motive, alienates one from one's friend (cf. Williams 1973 and Markovits 2010).

This version of the objection moves too quickly. To see this, consider the difference between Harmonious Act Consequentialism and Harmonious Eight-Ball, which maintains that one ought to do whatever the eight-ball randomly decides, and that one ought to be motivated to do it because the eight-ball said so. In the case of Harmonious Act Consequentialism, but not Harmonious Eight-Ball, facts about the benefits to one's friend are part of the *full* explanation of why one ought to benefit Joe, if not part of the *immediate* explanation for why one ought to do so. So facts about the benefit to Joe will be part of one's motivation in performing the action, alongside facts about the various other costs and benefits of doing so.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

This may not yet be fully satisfying. Every fact about the anticipatable effects of your alternatives on anything evaluatively significant until the end of time will show up in your motivational structure, *qua* part of the explanation of why you ought to perform this action. But another refinement is available to the Harmonious Act Consequentialist: they can acknowledge a distinctive level of normative explanation, between facts about values and facts about oughts, consisting of facts about reasons; and they might distinguish between reasons and their background conditions (cf. Schroeder 2007 and Maguire 2016). The facts about actions promoting states of affairs might motivate an agent on the condition that those states of affairs are valuable, and in proportion, other things being equal, to their value. The distinction between one's motivating reason and its background condition helps to ensure that some facts about the individual being helped feature in the content of specific motives; the proportionality condition elevates the significance of the impacts on the beneficiary above other parts of the full explanation. One might then add a pragmatic story, again along the lines sketched in (p. 405) Schroeder (2007), that explains why it is most appropriate to attend, in one's practical reasoning, to those reasons that are most likely to make a difference to what you ought to do. Hence, it will often be appropriate to attend, in one's reasoning, specifically to facts about benefiting friends and family (among others).

It may also happen that, in many cases, one has *more reason* to help one's friends than strangers, just on purely impartial grounds (cf. Jackson 1991). Take a case in which you are more motivated to provide the same resource to a friend than to a stranger in virtue of having better evidence about the contribution of the benefit to your friend's welfare. Your evidence suggests that the resource will have the same impact on their welfare. But you are justified in being more confident that the resource will make this impact on your friend than on the stranger (assume the only alternative to this positive welfare impact would be welfare neutral). This is a case in which your motivation to help your friend is stronger than your motivation to help the stranger. And, other things being equal, you will, in fact, help your friend. Continue to assume that the content of your motivation is not relevantly alienating. In this case, the *resultant* strength of your motivating reason is greater for your friend than for the stranger, for the same anticipated welfare impact.

So there is some sense in which the Harmonious Act Consequentialist can not only advocate being motivated distinctly by facts about one's friends, but also, in some cases, being more motivated by those facts than by facts about the impacts of one's actions on strangers. Is this latter fact about the resultant strength of your motive sufficient to properly recognize your friend in your practical reasoning? We suggest not. A friend might reasonably hope for more recognition than this: for their interests to play a more significant role in your practical reasoning (cf. Darwall 1977). This is clearest when we step back to consider the rule one is following in this case: that one ought to give someone's interests more weight in proportion to one's justified confidence about successfully helping them. This entails, in many other cases, that one will give the stranger's interests more weight than one's friends whenever it happens that one's evidence about the contri-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

bution to their welfare is slightly stronger. Intuitively, following this rule in one's practical reasoning is not sufficient to properly recognize one's friend as such.

Furthermore, although the motive content of the Harmonious Act Consequentialist does include a direct reference to their friend, the agent is motivated to help their friend *in virtue of* their friendship (or their associated concern for their friend) only insofar as facts about their relationship are relevant to the consequentialist calculation (e.g., "my friend's happiness will be increased more by my visit than by a stranger's visit"). In the agent's mind, the friendship has no normative or motivational force outside the role it plays in calculations about neutral value.

Those considerations apply to the case in which Joe is a good friend. But—as the history of the concept of alienation in political contexts also suggests—the alienation challenge does not merely afflict cases involving partiality. An agent might be alienated from a perfect stranger, to whom she has no special commitments or obligations, in virtue of her failure to participate in some ideal of community. Perhaps the proper adherence to norms manifesting mutual concern or recognition involves some nonmaximizing, or even some non-promotion-based, response (cf. Honneth 1992; Scanlon 2008; (p. 406) Cohen 2009; and Scheffler 2015). Common-sense accounts of virtues such as compassion, benevolence, and sympathy, and of social roles such as citizen, neighbor, and teacher, will also often require a less abstract, less calculative, or less impartial response than the Harmonious Act Consequentialist's (cf. Brennan and Pettit's discussion of virtues in their 1986; see also Blum's discussion of vocations and challenges to the personal/impersonal distinction in his 1993).

This introductory discussion suggests that, in doing something that benefits someone, none of the following is sufficient to avoid alienation: having the fact that you are benefitting the person as part of your motivation; being distinctly motivated by the fact that you will benefit the person; being saliently motivated by the fact that you are benefitting the person; or being more motivated by the benefit to the person than by the same benefit to another. The discussion also raises various questions. One concerns the *stringency* of your motivation: that you take your friend's interests to be more important than others in the right kind of way. One concerns the *robustness* of your motivation: that you take your friend's interests to be so important across the right range of circumstances. Another concerns the *source* of this robustness: this more modally robust orientation is based on a concern for one's friend as such. A further issue concerns the underlying *justification* for one's motivation, that it is permissible, all things considered, to be so motivated in matters concerning one's friend. Analogues of these points may also apply to actions helping strangers and to actions in fulfilment of projects. In the remainder we will consider four kinds of strategies employed by consequentialistically inclined theorists in response to these questions.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

3. Some Alternative Consequentialist Strategies

3.1. Hybrid Theories

According to Hybrid Theories, consequentialism is not supreme. The deontic facts explained by value promotion do not entail facts about what one *just plain* ought to do.[10] On some versions, a consequentialist principle fully explains the moral ought, but it is a further question how the moral ought interacts with other kinds of deontic oughts, in particular those arising from personal relationships and projects. On this view, personal relationships are beyond morality. Other versions don't distinguish a moral ought from the just plain ought. Instead, they hold that morality itself has a broadly consequentialist structure, but that it permits exceptions from value maximization for acts in the context of personal relationships and other commitments (see Scheffler 1982).

 The implications for these two versions of a Hybrid Theory are different.[11] On the "Distinct Oughts" view, the moral ought might be explained consequentialistically, perhaps by Act Consequentialism; it is just that morality isn't always overriding. The overall *ethical* theory includes other nonmoral sources of authoritative normativity. This view is structurally similar to W. D. Ross's deontological view (1930), with impersonal benevolence as one *prima facie* duty among others. This is a case in which alienation is avoided by the overall ethical theory only by the authority of fundamental nonconsequentialist principles. Furthermore, within its own jurisdiction, consequentialism still faces an alienation challenge. On the "Partial Option" view, morality itself permits responsiveness to partial considerations. No further appeal is made to distinct moral and all-things-considered oughts. On this view, it is not clear whether morality is alienating at all. However, again, it is significant that the title of the seminal defense of this position is *The Rejection of Consequentialism*.

It is also worth noting that both versions may also run into difficulties accounting for nonalienation in *non*partiality cases. On either version of the Hybrid Theory, the agent's motive in doing something that helps a stranger will be the same as that of the regular Harmonious Act Consequentialist (assuming there are no nonmoral oughts, or partial permissions, in play). As we have seen, it is an open question whether this motivational profile can fully account for the role of virtues, social roles, and civic relationships in our ethical life.

3.2. Relative Value Theories

According to Relative Value Theories, what one ought to do is explained, not by what would promote the most neutral value, but by what would promote the most "relative value," that is, the most value relative to the agent.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

There are various ways to understand the notion of "value relative to the agent."[12] One is to start with some idealization of the agent's desires, or with some fundamental notion of what an agent has reason to desire, or with what it would be fitting for the agent to desire (see, e.g., Chappell 2019 and Portmore 2014). Such facts may then be used to explain what an agent ought to do and how they ought to be motivated (if motives and desires are distinct). Our central concern with such approaches is that they abandon the key consequentialist idea altogether, which is that deontic facts are explained by the agent's relation to neutral value.

An alternative view starts with facts about neutral value and adds the idea of modification: these values are ranked by facts about their "moral distance" from the agent. The function from the neutral values and their moral distance from the agent yields a set of "relative value" facts. These are then used to explain the deontic facts (for versions of this gambit, see Hurka 2001; Maguire 2013; Bader 2016; and Maguire 2017).

(p. 408) How would this apply to the ethics of motives? Harmonious Relative Value Theories will maintain that one ought to be motivated to promote states of affairs that are relatively valuable, in proportion to their relative value. To (over)simplify matters, assume, on the one hand, that friendship calls for responsiveness just to welfare facts, and, on the other hand, that all and only welfare facts are neutrally valuable. Then the *facts* that motivate one will be the same whether one is responding just as a friend, or just as a consequentialist, or as a Harmonious Relative Value Theorist. What about the *weight* of these motives? Does the relative value of some benefit to a friend correspond to a motive of the same strength as might be demanded by some plausible normative ideal of friendship? This will turn on how particular theories characterize "moral distance." But this concept is designed, more or less, to yield the correct result here. So far, then, there is no obvious alienation in cases of partiality.

It is worth looking closely at the difference between the resultant motivational profile of the Harmonious Jackson-style Decision Theoretic Consequentialist and that of the Harmonious Relative Value Theorist. The latter is motivationally sensitive to more than agent-neutral facts about the promotion of welfare, relative to available evidence. In addition, they are motivationally sensitive to whatever it is that accounts for "moral distance." Suppose, for a moment, that one is "morally closer" to one patient than another to the extent that one actually cares more about the one than the other. Then the strength of one's motivation to help someone one cares more about than another will be explained partly by the facts about the neutral value of welfare and also partly by the fact that one cares more about the one than the other. That seems like just the kind of motivational sensitivity that might avoid alienation.

Furthermore, in this case, as contrasted, in particular, with the "Distinct Oughts" version of a Hybrid Theory, alienation is not avoided by one principle *in spite of* the alienating consequentialist moral principle. Rather the consequentialist principle is playing a partial role in the full explanation of the all-things-considered ought, alongside moral distance.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

And, indeed, it seems proper for a friend to be responsive to the (neutrally valuable) facts about the impacts on her friend's welfare of this or that intervention.

But there is a similar concern here about what in this ethical theory secures nonalienation. The concern is that it is not exactly the *consequentialist* principle that gets the motive weight right, but whatever explains one's "moral distance" from the patient. The notion of moral distance is not coming from the consequentialism but from some other principle: perhaps strength of care, as in our example, or perhaps simple egoism. Whatever plays this role has fundamental normative authority.

One possibility for the Relative Value Theorist is to offer a consequentialist account of "moral distance" (cf. Maguire 2016). For it is rather implausible that actual facts about an agent's psychology are fully adequate to play the role of a normatively authoritative modifier. One often fails to care as one should. Perhaps, then, moral distance can be explained by what one *ought* to care about; this "ought" might, in turn, admit of a simple consequentialist account: have whichever caring profile would be best. It is hard to believe Frank Jackson that, case by case, more value will be promoted by helping the near and dear. But it is more plausible—particularly in more egalitarian worlds—that (p. 409) having a caring profile according to which one cares more about one's near and dear than strangers will promote more value than a caring profile that ranks everyone equally (cf. Maguire 2016).

However, we are doubtful whether such a position would yield fully satisfying extensional results concerning actions and motives. But even if so, this position strikes us as unnecessarily baroque. It is the conjunction of an Act Consequentialist theory of the existence of reasons with a combination of an Act Consequentialist and a Leveled Consequentialist theory of the weight of reasons. At this point, a consequentialist would be advised to opt instead for a full-blooded Leveled Consequentialism, against which we will shortly present a range of challenges.

3.3. Global Consequentialism

Global Consequentialism is the universalized version of direct consequentialism.[13] According to Global Consequentialism, the most value-promoting alternative for every given "focal point" (actions, motives, forms of government, roof colors, etc.) ought to obtain.[14]

Global Consequentialism is often presented as a solution to the problems that face Harmonious Act Consequentialism. The latter often gives the wrong answer to the question of what ought to motivate an agent in interpersonal contexts (both when the agent is interacting with the near and dear and with strangers) and, plausibly, in certain intrapersonal contexts, as when an agent is pursuing a ground project (on which more shortly). But in many cases, having the "nonalienated" motive in such contexts promotes more value than not. This permits a conjecture: direct consequentialism about motives might have a more intuitively attractive set of deontic prescriptions for motives (cf. Sidgwick 1874; Adams 1976). The Global Consequentialist is unwilling to relinquish the direct assessment of

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

acts, and instead relinquishes the idea that there is any direct normative relationship between the assessment of acts and the assessment of motives. For that matter, the view also relinquishes any direct normative relationship between ways of reasoning and actions, between commitments and ways of reasoning, and so on.

As far as alienation is concerned, there are various problems with Global Consequentialism.[15] The first is this. It may be true that the motives, commitments, modes of reasoning, and so on advocated by some ideal of friendship are consequentialistically better than those advocated by Harmonious Act Consequentialism. But it does (p. 410) not follow that the former are the *best* available—not even taking into account the disvalue, if there is any, of alienation itself. Presumably it will often be consequentialistically better to be motivated to abandon our friends and family in a cocaine-fueled pursuit of hedge-fund-managed millions, to be gifted to the poor.

The second problem is this. Take a case in which, according to Global Consequentialism, you ought to be most motivated to *x* and yet in which you ought not *x*. Lots of familiar kinds of partiality cases will be like this: you ought to be motivated to help your friend or your child, but you ought to help the slightly needier stranger. The problem is that, in many such cases, the very normative ideal that provides support for the claim that you ought to be motivated to help your friend rather than the stranger also provides support for the claim that you ought to help your friend rather than the stranger; and that you ought to help your friend *out of* your greater concern for your friend. Normative ideals can call for partial rather than impartial behavior. But since Global Consequentialism entails Act Consequentialism, it entails that one ought always to do whatever is impartially best.

A closely related problem concerns the intrapersonal ideal of integrity. According to Global Consequentialism, the fact that one ought to have some commitment that would make it fitting to be motivated to *x* in S is explanatorily irrelevant to whether one ought to be so motivated, since being committed and being motivated are distinct focal points.[16] Suppose that one ought not to be so motivated. A suitably apprised Global Consequentialist will find herself struggling to satisfy these two directives. If she has the motives she ought to have, she will fail to have the motives that fit her commitments—in contravention of a plausible normative ideal. Commitments can survive occasional exceptions. But it is not so clear that they can survive the sober, clear-headed judgement that they have no motivational significance, now or ever. So, having the Global Consequentialist motive also undermines one's claim to having the relevant commitment at all.

One way to sum this up is to point out that nonalienation relative to a range of prominent normative ideals calls for *multifocal fittingness:* for an agent to be oriented toward another person or project in her actions, motives, affects, and ways of thinking, where these different states are all normatively integrated at a time and across time. In disaggregating these different responses, Global Consequentialism gives precisely the wrong kind of prescription.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

4. Leveled Consequentialism

4.1. Characterization of Leveled Consequentialism

The discussions so far have variously pointed in the direction of a version of consequentialism that directly assesses higher-level psychological states, which in turn yield more (p. 411) specific agential attitudes. Call such theories Leveled Consequentialisms (see, e.g., Rawls 1955; Railton 1984; Pettit and Brennan 1986; Tännsjö 1995; Norcross 1997; Mason 1998; Mason 1999; Tedesco 2006). Examples of such higher-level states include commitments, values, identities, decision procedures, roles, caring profiles, character traits, and, simply, dispositions. These psychological structures are robust in the sense that they tend to persist over time and possibilities. They are higher-level in the sense that they explain—causally, by making fitting, or both—a range of more specific, lower-level states such as motives, thought patterns, and actions. The Leveled Consequentialist has a "two-tier" psychological structure, and these two tiers are potentially responsive to different kinds of justifications. The higher-level states generate, coordinate, and support a set of motives, thought patterns, and/or behavioral tendencies. The higher-level states are responsive, in their turn, to consequentialist considerations.

There are both indirect and global versions of Leveled Consequentialism, differing as to whether they maintain that one ought to act in response to one's higher-order states, or whether one ought nonetheless to act in whatever way would be best.[17] The latter is simply part of Global Consequentialism, maintaining that you ought to do whatever is best and that you ought to be committed however is best, even though the two might conflict (Hare 1981; Railton 1984; Pettit and Brennan 1986). Indirect Leveled Consequentialism maintains that you ought to do whatever you ought to be committed to (or disposed to, or what have you) (Rawls 1955; Mason 1998; Tedesco 2006).

We'll put aside Global Leveled Consequentialisms, and use Leveled Consequentialism to refer just to Indirect Leveled Consequentialisms. Notice that Leveled Consequentialisms need not be "self-effacing."[18] To see this, stick with Leveled Consequentialisms that concern higher-order states that involve principles of reasoning. Some such theories maintain that one ought to employ some nonconsequentialist decision procedure in certain first-order contexts. But they also maintain that the ethical facts about actions are themselves responsive to such nonconsequentialist considerations in those contexts. Such theories do tell one to employ a consequentialist decision procedure in higher-level matters, concerning whether to accept these first-order nonconsequentialist principles, but again the ethical facts are responsive to consequentialism on that level. So there is congruence at both levels between the criterion of rightness and the right decision-making procedure. As before, however, notice that this comes at the "cost" of relinquishing explanatory autonomy to nonconsequentialist considerations.

Leveled Consequentialisms strive to take seriously the idea that some of the most important goods in life consist in proper participation in certain normative ideals, such as those of familial love and caring friendships. These theories aim to preserve the fact that such

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

participation can require not just behavior but motives, affects, particular ways of reasoning, and, in general, commitments to the activity for its own sake. The two-tier structure attempts to uphold the autonomy of standards of (p. 412) participation, while restricting fundamental authority to the consequentialist assessment of participation itself.

We wish to raise three kinds of concerns about the leveled approach. The first concerns its feasibility: can the distinction between levels be upheld? A second concerns its applicability: does the alienation objection also apply at the higher level? What conditions are acceptable on the commitment of a friend, citizen, and so on? The third concerns integrity: are Leveled Consequentialists thereby alienated from their own particular commitments, or types of commitments, or their relation to themselves as agents with commitments?

4.1.1. The Collapse Objection

According to the Collapse Objection, the distinction between two levels does not enable the agent to avoid alienation, for the fact that the higher level admits of consequentialist justification unavoidably infects the standing or nature of the lower-level states. To take a specific version of the objection, one is not *really* doing something for one's friend because of one's friendship, but because the friendship is supported by the most value promoting disposition, commitment, or whatever.

We are optimistic that this objection can be avoided. The structuring of practical thought into distinct levels, spheres, domains, and so on seems to us a deep and unavoidable fact of life. As Rawls (1955) argued, we frequently sustain our commitment to one set of rules, which provide their own set of justifications, on the basis of our commitment to some other kind of justification. For instance, we might accept a practice of punishment for wrongdoing because of its utilitarian advantages, but punish independently of the utilitarian advantages of doing so. To take some further examples, we are fluent in transitions from "I" reasoning to "we" reasoning in everyday life; we can also reason differently *qua* teacher, *qua* parent, and *qua* friend. There is a mixture of empirical, theoretical, and normative issues here that we cannot sort out just now (for recent work on associated empirical issues see Seligman et al. 2017). Suffice it to say that interesting work is required to be done in distinguishing the cases in which levels do collapse, of which Rawls' "summary rules" are an example, from those in which they do not, of which Rawls's "practice rules" are an example.

4.1.2. The Higher-Level Alienation Objection

Earlier we encountered the plausible conjecture that nonalienation requires not just particular motives, but also that those motives have a particular source: that one is motivated to do things out of the right kind of commitment or other higher-level state. Here is a related concern. It might be argued that certain kinds of reasons for having friendly commitments can themselves undermine the status of the relationship as a friendship (Kapur 1991; Cocking and Oakley 1995). The issue here doesn't concern the status of the first-or-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

der motives. It may be that the collapse can be avoided. But it is charged that the higher-order motives themselves are sufficient to vitiate the normative ideal.

To take an important example, Cocking and Oakley argue that the ideal of friendship dictates certain conditions under which one would enter into and leave a friendship—the (p. 413) "governing conditions" of one's commitment to the relationship. They argue that the responsiveness of higher-level states to consequentialist considerations would mean that one would fail to meet these "governing conditions" for one important ideal of friendship, and hence that one would fail to be a friend at all. Indeed, they maintain that the governing conditions are "most significantly responsible for the problem of alienation" (1995, 111).

Presumably, certain kinds of conditionality are compatible with friendship. You might be disposed to relinquish your commitment if you discover that your friend is a serial killer, or if the friend has an affair with your spouse, or simply if your relationship ceases to be enjoyable for you both. That conditionality doesn't interfere with the quality of your relationship if there have been no murders or affairs. But Cocking and Oakley argue that straightforward conditionality of a consequentialist kind would be incompatible with friendship.

By way of analogy, they discuss the example of an ambitious philosopher befriending a famous professor for professional advancement. However, the ambitious philosopher has successfully ensured that only his commitment to the friendship is responsive to his professional goals; his motives are not counterfactually responsive in this way. The individual actions of the ambitious philosopher are not motivated by professional ambition, but the philosopher is disposed to relinquish the relationship if it no longer serves his ambition.

We agree that there is alienation in the professor case, relative to a familiar normative ideal of friendship. However, the alienation feels a little different to us. Assume the Collapse Objection can be avoided, and that the ambitious philosopher really does *care* about the famous professor. Cocking and Oakley allow that this commitment may not be dynamically responsive to the philosopher's judgement that the friendship no longer serves his ambitious interests. The philosopher might need to work hard to get himself to stop caring so much about the famous professor, perhaps to focus on networking with the deans instead. Granted, something is wrong with his relationship with the professor. But something is right about it, too. The professor would have been *more* alienated had the philosopher's ambition led him to try to gainsay his opinions at every turn, for instance.

An option for the consequentialist at this point is to add another level. Perhaps it is not individual relationships themselves that are assessed for value maximization, but the governing structures that coordinate and support individual relationships (Mason 1998; Tedesco 2006). But in reply, one might modify Cocking and Oakley's objection and argue that it is alienating for a Leveled Consequentialist's friendships to be conditional upon a higher-level governing structure, which one accepts if and only if it is appropriately value promoting. But at this point, while there is clearly something distasteful about the ambitious philosopher's character, it is less clear that the ambitious philosopher would be

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

alienated from the famous professor. Imagine if all these facts were common knowledge, and the famous professor accepted the higher-level governing structure on the same basis. Perhaps the famous professor is friends with the ambitious philosopher only on the basis of a general commitment to having others thinks he is (p. 414) generous and gregarious. But they both really do care about each other. This pair sounds vicious but not necessarily alienated—or as alienated—from each other.

When we turn from conditionality on egoistic considerations, to conditionality on consequentialist considerations, things are even murkier. For it is plausible enough that *some* higher-order conditionality on purely extrinsic moral matters is permitted by familiar ideals of friendship. Suppose that you are an epidemiologist with specialized training in the treatment of some rare disease in the days before long-distance communication. You might be friends with someone, even though you both know that you might have to move overseas indefinitely and at short notice. There would be certain kinds of investments you would both be less likely to make in the friendship. You won't plan to buy a house together, for instance. So there may be degrees of commitment that you are prevented from participating in, and you might be alienated from ideals involving the prospect of such a degree of commitment. But you might have a valuable committed friendship all the same, with most of your first-order responses intact. You see each other a couple of times a month to enjoy shared activities; you help each other out occasionally; you check in when one or the other has something to celebrate or commiserate about; and so on.[19] As this conditionality ascends to higher levels in one's commitment structure, it likewise sounds less problematic. Perhaps one is disposed to stop having any close friends at all, if circumstances change enough that having friends can no longer be justified on consequentialist grounds.

However, notice that this correlates with consequentialism itself playing an increasingly diminished role in the overall organization of one's life. As the consequentialist assessment ascends to higher levels, one is responsive to more and more commitments, each with its own nonconsequentialist authority.

4.1.3. The Moral Self Objection

A further question remains, namely whether one will be alienated from *oneself* by one's commitment to Leveled Consequentialism.[20] A certain amount of alienation from our actual commitments is a necessary condition on being able to reflect critically on one's own, and our collective, life (cf. Railton 1984). It is not problematic for an ethical theory to require the egoist or militant racist or perhaps even the masochist to modify or abandon a commitment, project, role, relationship, or personality trait. And perhaps the true ethical theory will require most of us to adopt more of the sorts of commitments consequentialists talk about, for instance to donating more to organizations helping people in extreme poverty, resisting the factory farming system, and caring more about future generations, all at a trade-off to our current lifestyles and resource allocations. But all this leaves open the questions of whether the true ethical theory subjects our identities, ground projects, and so on to *consequentialist* assessment, and whether an (p. 415) ethical theory that does so is alienating. Insofar as consequentialism evaluates one's higher-order states just in

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

terms of the neutral value of their total consequences, people will likely be unable to honor a range of deeply held, and intuitively reasonable, ethical and nonethical commitments. The key question here is: What standards of ethical assessment can one accept without somehow alienating oneself?

The challenge for consequentialism is perhaps clearest in cases where value differences are marginal. Consider a quantitatively skilled public policy professional working on US education who, after some careful research, expected value calculations, and considerations of replaceability in the job market concludes that the expected impact of her career would be slightly higher if she switched to investment banking and donated much of her earnings to the best nonprofit working on education. Imagine further that this person is fascinated by public policy and political science, but finds financial analysis relatively unfulfilling. It seems alienating for an ethical theory to require this person to give up her fulfilling career in policy for an unfulfilling career in banking simply because of a marginally higher expected impact. To take another example, imagine someone whose chief hobby in life is chess. This person realizes that if he gave up chess and devoted the time and energy to gardening instead, he would generate more neutral value, since people could aesthetically appreciate the resulting garden. Again, it seems alienating for an ethical theory to require this person to renounce chess in favor of gardening, simply because the change of project would promote somewhat more neutral value.

As pointed out in the discussion of Global Consequentialism, it doesn't follow from the fact that a nonalienating set of commitments would add value that the nonalienating set would be consequentialistically *best*; nor that even if it were, consequentialism provides the intuitively correct justification for the commitments. Likewise, it does not follow from the evident fact that many peoples' commitments are morally dubious, and even that they are insufficiently sensitive to facts about neutral value promotion, that the proper standard of critical assessment is some version of Supreme Consequentialism. Particular commitments may involve sensitivity to neutral value in their own way. For instance, caring about a friend involves some sensitivity to what is good for the friend, which is (at least conditionally) of value. Likewise committing to being a good doctor involves positive sensitivity to values. And one might have a larger commitment to regulating one's commitments in part by their social usefulness, measured roughly consequentialistically. But this falls far short of finding it acceptable to assess one's commitments wholesale just in terms of their consequential value.

One final point. For all we have said, someone, or some society, might come to accept some version of Supreme Consequentialism on the basis of authentic ethical reflection. They might also accept compatible normative ideals concerning the nature of the self, and of society, and of the very enterprise of morality. But this kind of coherence is not sufficient for nonalienation. On the contrary, this would just make them *more* alienated relative to alternative and more attractive ideals (see Cohen 1968, 215, on "ersatz humanity").

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(p. 416) 5. Conclusion

We considered four consequentialist strategies for avoiding alienation objections of various kinds. None is fully satisfying.

One option—Hybrid Theory—is to restrict one's consequentialism to a subordinated role in one's overall theory of what one ought to do. One worry with this view is whether the residual consequentialism will itself be alienating, offering too abstract a standard even in impartial cases. The more general concern is that the overall theory of what one ought to do is not consequentialist, and alienation is avoided just by the nonconsequentialist part. This seems less a victory for consequentialism than a defeat.

Another option—Relative Value Theory—is to mix one's consequentialist assessment of options with an agent-relative ranking, using the resulting agent-relative ranking to explain facts about what one ought to do. The central worry with this approach is that, again, insofar as alienation is avoided, it is by the nonconsequentialist constituent in the proposal, namely the grounds for whatever yields the agent-relative ranking.

A third option—Global Consequentialism—is to disaggregate one's assessment of motives and actions. Plausibly enough, an impartial set of motives will promote less value than a partial set of motives; partial motives can then be advocated on the grounds of neutral value promotion. Problems will persist when nonalienation intuitively requires not just partial motives but partial actions. Further problems arise upon consideration of the agent in conformity with Global Consequentialism—their values, commitments, motives, and actions all askew.

In these cases, however, progress seems to be made in ethical theory, if not exactly toward a consequentialist ethical theory, for we have a better understanding of how we might integrate values and commitments with different kinds of ethical assessment.

The fourth option—Leveled Consequentialism—advocates compliance with higher-level states, such as one's commitments. It is plausible that a consequentialist assessment of such higher-level states will advocate commitments to oneself and others in ways that involve less alienation, at least, than traditional Act Consequentialism. Or perhaps one can restrict consequentialism to even higher states that govern one's selection of commitments, or perhaps of social, political, and economic institutions and practices within which one lives one's life. At each greater height, consequentialism cedes more normative authority to nonconsequentialist rules and institutions. But it seems also to be alienating to make certain trade-offs, and to make or reject certain commitments, on purely consequentialist grounds, even at these lofty theoretical heights.

It remains an open question whether consequentialism will play some more limited explanatory role in a fully satisfying ethical theory, or whether the considerations that speak in its favor—facts about proper sensitivity to value, the appropriateness of dominance reasoning or maximization in different domains, and so on—will be fundamentally ex-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

plained in other ways, which may be more straightforwardly reconciled with plausible normative ideals.

References

Adams, R. M. 1976. “Motive Utilitarianism.” *The Journal of Philosophy* 73, no. 14: 467–481. **https://doi.org/10.2307/2025783**.

Arpaly, Nomy. 2011. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford: Oxford University Press.

Bader, Ralf. 2016. “Conditions, Modifiers, and Holism.” In *Weighing Reasons*, edited by Errol Lord and Barry Maguire, 27–55. New York: Oxford University Press.

Blum, Lawrence. 1993. “Vocation, Friendship, and Community: Limitations of the Personal-Impersonal Framework.” In *Identity, Character, and Morality*, edited by Owen Flanagan, 173–198. Cambridge, MA: MIT Press.

Brudney, Daniel. 2014. “The Young Marx and the Middle-Aged Rawls.” In *A Companion to Rawls*, edited by Jon Mandle and David Reidy, 450–471. Malden, MA: Wiley Blackwell.

Chappell, Richard Yetter. 2019. “Fittingness Objections to Consequentialism.” In *Consequentialism: New Directions, New Problems*, edited by Christian Seidel, 90–112. New York: Oxford University Press.

Cocking, Dean, and Justin Oakley. 1995. “Indirect Consequentialism, Friendship, and the Problem of Alienation.” *Ethics* 106, no. 1: 86–111. **https://doi.org/10.1086/293779**.

Cohen, G. A. 1968. “Bourgeois and Proletarians.” *Journal of the History of Ideas* 29, no. 2: 211–230. **https://doi.org/10.2307/2708577**.

Cohen, G. A. 2009. *Why Not Socialism?* Princeton, NJ: Princeton University Press.

Cox, Damian. 2005. “Integrity, Commitment, and Indirect Consequentialism.” *The Journal of Value Inquiry* 39, no. 1: 61–73. **https://doi.org/10.1007/s10790-006-1571-7**.

Crisp, Roger. 1992. “Utilitarianism and the Life of Virtue.” *The Philosophical Quarterly* 42, no. 167: 139–160. **https://doi.org/10.2307/2220212**.

Dancy, Jonathan. 2019. *Practical Shape*. Oxford: Oxford University Press.

Darwall, Stephen. 1977. “Two Kinds of Respect.” *Ethics* 88, no. 1: 36–49. **https://doi.org/10.1086/292054**.

Driver, Julia. 2014. “Global Utilitarianism.” In *The Cambridge Companion to Utilitarianism*, edited by Ben Eggleston and Dale E. Miller, 166–176. New York: Cambridge University Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Feldman, Fred. 1993. “On the Consistency of Act- and Motive-Utilitarianism: A Reply to Robert Adams.” *Philosophical Studies* 70, no. 2: 201–212. **https://doi.org/10.1007/bf00989590**.

Greaves, Hilary. 2020. “Global Consequentialism.” In *The Oxford Handbook of Consequentialism*, edited by Douglas W. Portmore. Oxford: Oxford University Press.

Hare, R. M. 1981. *Moral Thinking: Its Levels, Method, and Point*. Oxford: Oxford University Press.

Honneth, Axel. 1992. *The Struggle for Recognition*. Cambridge, MA: MIT Press.

Hooker, Brad. 2000. *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*. Oxford: Clarendon Press.

Hurka, Thomas. 2001. *Virtue, Vice, and Value*. New York: Oxford University Press.

Jackson, Frank. 1991. “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection.” *Ethics* 101, no. 3: 461–482. **https://doi.org/10.1086/293312**.

Jaeggi, Rahel. 2014. *Alienation*. Translated by Frederick Neuhouser and Alan E. Smith, edited by Frederick Neuhouser. New York: Columbia University Press.

Kagan, Shelly. 2000. “Evaluative Focal Points.” In *Morality, Rules and Consequences: A Critical Reader*, edited by Brad Hooker, Elinor Mason, and Dale E. Miller, 134–155. Edinburgh: Rowman & Littlefield.

(p. 418) Kain, Philip. 1982. *Schiller, Hegel, and Marx*. Kingston: McGill Queens University Press.

Kapur, Neera Badhwar. 1991. “Why It Is Wrong to Be Always Guided by the Best: Consequentialism and Friendship.” *Ethics* 101, no. 3: 483–504. **https://doi.org/10.1086/293313**.

Kaufman, Walter. 1970. Introduction to *Alienation*, by Richard Schacht. Garden City, NY: Doubleday.

Leopold, David. 2018. “Alienation.” In *The Stanford Encyclopedia of Philosophy* (Fall 2018 edition), edited by Edward N. Zalta. **https://plato.stanford.edu/archives/fall2018/entries/alienation**.

Louise, Jennie. 2006. “Right Motive, Wrong Action: Direct Consequentialism and Evaluative Conflict.” *Ethical Theory and Moral Practice* 9, no. 1: 65–85. **https://doi.org/10.1007/s10677-005-9000-8**.

MacIntyre, Alasdair C. 1984. *After Virtue: A Study in Moral Theory*. Notre Dame, IN: University of Notre Dame Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Maguire, Barry. 2013. “Values, Reasons, and Ought.” PhD dissertation, Princeton University.

Maguire, Barry. 2015. “Grounding the Autonomy of Ethics.” In *Oxford Studies in Metaethics*: *Volume 10*, edited by Russ Shafer-Landau, 188–215. Oxford: Oxford University Press.

Maguire, Barry. 2016. “The Value-Based Theory of Reasons.” *Ergo* 9, no. 3: 233–262. **http://dx.doi.org/10.3998/ergo.12405314.0003.009**.

Maguire, Barry. 2017. “Love in the Time of Consequentialism.” *Nous* 51, no. 4: 686–712. **https://doi.org/10.1111/nous.12169**.

Maguire, Barry, and Woods, Jack. Forthcoming. “The Game of Belief.” *The Philosophical Review.*

Markovits, Julia. 2010. “Acting for the Right Reasons.” *The Philosophical Review* 119, no. 2: 201–242. **https://doi.org/10.1215/00318108-2009-037**.

Mason, Elinor. 1998. “Can an Indirect Consequentialist Be a Real Friend?” *Ethics* 108, no. 2: 386–393. **https://doi.org/10.1086/233810**.

Mason, Elinor. 1999. “Do Consequentialists Have One Thought Too Many?” *Ethical Theory and Moral Practice* 2, no. 3: 243–261. **https://doi.org/10.1023/A:1009998927955**.

Mason, Elinor. 2002. “Against Blameless Wrongdoing.” *Ethical Theory and Moral Practice* 5, no. 3: 287–303. **https://doi.org/10.1023/A:1019671210369**.

McLeod, Owen. 2001. “Just Plain “Ought.’” *The Journal of Ethics* 5, no. 4: 269–291. **https://doi.org/10.1023/A:1013934513554**.

Norcross, Alastair. 1997. “Consequentialism and Commitment.” *Pacific Philosophical Quarterly* 78, no. 4: 380–403. **https://doi.org/10.1111/1468–0114.00045**.

Oakley, Justin, and Cocking, Dean. 2001. *Virtue Ethics and Professional Roles*. Cambridge: Cambridge University Press.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.

Pettit, Philip, and Brennan, Geoffrey. 1986. “Restrictive Consequentialism.” *Australasian Journal of Philosophy* 64, no. 4: 438–455. **https://doi.org/10.1080/00048408612342631**.

Pettit, Philip, and Smith, Michael. 2000. “Global Consequentialism.” In *Morality, Rules and Consequences: A Critical Reader*, edited by Brad Hooker, Elinor Mason, and Dale E. Miller, 121–133. Edinburgh: Rowman & Littlefield.

Portmore, Douglas W. 2014. *Commonsense Consequentialism: Wherein Morality Meets Rationality*. Oxford: Oxford University Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Railton, Peter. 1984. “Alienation, Consequentialism, and the Demands of Morality.” *Philosophy and Public Affairs* 13, no. 2: 134–171. **https://www.jstor.org/stable/2265273**.

Rawls, John. 1955. “Two Concepts of Rules.” *The Philosophical Review* 64, no. 1: 3–32. **https://doi.org/10.2307/2182230**.

(p. 419) Raz, Joseph. 1986. *The Morality of Freedom*. Oxford: Oxford University Press.

Ross, W. D. 1930. *The Right and the Good*. Oxford: Clarendon Press.

Scanlon, T. M. 2008. *Moral Dimensions*. Cambridge, MA: Harvard University Press.

Schacht, Richard. 1970. *Alienation*. Garden City, NY: Doubleday.

Scheffler, Samuel. 1982. *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*. Oxford: Clarendon Press.

Scheffler, Samuel. 2010. *Equality and Tradition*. Oxford: Oxford University Press.

Scheffler, Samuel. 2015. “The Practice of Equality.” In *Social Equality: Essays on What It Means to Be Equals*, edited by Carina Fourie, Fabian Schuppert, and Ivo Wallimann-Helmer, 21–44. Oxford: Oxford University Press.

Schroeder, Mark. 2007. “Teleology, Agent-Relative Value, and “Good.’” *Ethics* 117, no. 2: 265–295. **https://doi.org/10.1086/511662**.

Seligman, Martin E. P., Peter Railton, Roy F. Baumeister, and Chandra Sripada. 2017. *Homo Prospectus*. Oxford: Oxford University Press.

Sidgwick, Henry. (1874) 1981. *The Methods of Ethics*. Indianapolis: Hackett.

Stocker, Michael. 1976. “The Schizophrenia of Modern Ethical Theories.” *Journal of Philosophy* 73, no. 14: 453–466. **https://doi.org/10.2307/2025782**.

Tännsjö, Torbjörn. 1995. “Blameless Wrongdoing.” *Ethics* 106, no. 1: 120–127. **https://doi.org/10.1086/293781**.

Tedesco, Matthew. 2006. “Indirect Consequentialism, Suboptimality, and Friendship.” *Pacific Philosophical Quarterly* 87, no. 4: 567–577. **https://doi.org/10.1111/j.1468–0114.2006.00275.x**.

Way, Jonathan. 2017. “Reasons as Premises of Good Reasoning.” *Pacific Philosophical Quarterly* 98, no. 2: 251–270. **https://doi.org/10.1111/papq.12135**.

Whiting, Daniel. 2017. “Against Second-Order Reasons.” *Nous* 51, no. 2: 398–420. **https://doi.org/10.1111/nous.12138**.

Wilcox, William H. 1987. “Egoists, Consequentialists, and Their Friends.” *Philosophy & Public Affairs* 16, no. 1: 73–84. **http://www.jstor.org/stable/2265206**.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Williams, Bernard. 1973. “A Critique of Utilitarianism.” In *Utilitarianism: For and Against*, by Bernard Williams and J. J. C. Smart, 77–150. New York: Cambridge University Press.

Williams, Bernard. 1981a. “Persons, Character, and Morality.” In *Moral Luck: Philosophical Papers, 1973–1980*, by Bernard Williams, 1–19. Cambridge: Cambridge University Press.

Williams, Bernard. 1981b. “Utilitarianism and Moral Self-Indulgence.” In *Moral Luck: Philosophical Papers, 1973–1980*, by Bernard Williams, 40–53. Cambridge: Cambridge University Press.

Williams, Bernard. 1985. *Ethics and the Limits of Philosophy*. Cambridge, MA: Harvard University Press.

Wood, Allen. 1999. *Karl Marx*. London: Routledge.

Notes:

(1) Many thanks to Tristram McPherson, Austen McDougal, Milan Mossé, Oded Na’aman, Doug Portmore, and Jack Woods for excellent comments on earlier drafts.

(2) For historical discussions of alienation, see Schacht (1970) and Kain (1982). For recent discussions see Wood (1999); Brudney (2014); Jaeggi (2014); and Leopold (2018).

(3) In 1844, Marx discussed alienation from other people, but also from one’s labor, the product of one’s labor, and one’s species being. Hegel, Marx, and Fromm talked about alienation from the natural world. One might also wonder about cases of being alienated from activities, institutions, cultures, and so on. We will be principally concerned with alienation from oneself or others, more or less directly.

(4) On authoritative and nonauthoritative normativity, see Maguire and Woods (forthcoming) and references therein.

(5) Note that it is also an open question whether a given normative ideal is *operative* in a domain. We assume that an ideal is operative in a context only if a set of ideal-sustaining norms are accepted in the context. You might be alienated from a friend *by* the suboptimal norms of friendship that prevail in your social context, or by something else in spite of the availability of excellent friendship norms. There might be multiple norms of friendship available in a large social context, such that one would count as alienated relative to one but not another.

(6) We title an ethical *theory,* as opposed to an ethical principle that is a constituent in an ethical theory, as “consequentialist” only if it maintains that some consequentialist principle is supreme.

(7) As Julia Driver points out (Chapter 24, this volume), “indirect consequentialism” is used in two different ways in the literature. One is to denote an indirect deontic standard, as described in the main text. But it is sometimes used to distinguish consequentialisms

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

employing a sufficiently robust distinction between the deontic standard itself and the correct decision-making procedure. We'll stick with "indirect" for the former and use "self-effacing" (allowing that this comes in degrees) for the latter.

([8]) This needs to be refined. It is presumably compatible with plausible ideals of friendship that, for instance, if you happen to be teaching a close friend and giving him a good grade (knowing this will benefit the friend considerably), you do *not* do so because it will benefit him, but because the work merits it.

([9]) Alienation (as we are using the term) essentially involves the inhibition of states of an agent. Hence only things that can inhibit states of an agent can alienate. On the assumption that ethical principles are abstract objects, the unknowable truth of an ethical theory that no one accepts could not be alienating. (Although knowing that the ethical truth is unknowable might be.) It is individuals or collectives accepting (or believing, committing to, etc.) ethical principles that is inhibiting. An ethical theory is alienating only if is accepted (believed, committed to, etc.) either by oneself or others. In the latter case, social acceptance of some ethical principle may be manifest in patterns of behavior, social norms, or institutions that are alienating for particular individuals—or perhaps for everyone. In the former case, by accepting an ethical theory, one is oneself inhibited from participation in some kind of normative ideal, perhaps of integrity. This marks one important difference between alienation in the current context and the more familiar context of political economy. In the latter case, the socioeconomic systems that alienate are themselves the inhibitors; in the former case, ethical theories don't really alienate; acceptance of them (by oneself or others) does so.

([10]) For a discussion of the "all things considered" or "just plain" ought, see McLeod (2001) and Maguire and Woods (forthcoming).

([11]) Many thanks here due to Tristram McPherson.

([12]) For concerns about the very idea, see Schroeder (2007).

([13]) For general discussions and overviews of global consequentialism, see Driver (2014) and Greaves (Chapter 22, this volume). For defenses of the view, see Parfit (1984); Kagan (2000); Pettit and Smith (2000); Feldman (1993), and Greaves (Chapter 22, this volume).

([14]) Global Consequentialism is, of course, more complex in its fully-specified form. In particular, it is an open question which focal points are subject to deontic, rather than simply axiological, evaluation (see Greaves (Chapter 22, this volume) for further discussion). These considerations, while important, are mostly peripheral to our discussion, which focuses on focal points such as actions, motives, and commitments that are plausibly subject to deontic assessment.

([15]) Special thanks to Austen McDougal for discussion.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([16]) An interesting and concerning feature of Global Consequentialism, which has not been much remarked upon (though see Louise 2006, 82–84), is that the putatively nonethical individuation of focal points comes to carry a great deal of ethical weight.

([17]) Pity the fool who attempts to read this distinction back into the literature.

([18]) See fn. 7 for the distinction between indirect and self-effacing.

([19]) It is easy to construct examples in which the friendship was initiated for the wrong kind of reasons, as well.

([20]) Cf. Williams (1973); Williams (1981a); and Williams (1981b).

Calvin C. Baker

Calvin C. Baker is a PhD student in philosophy at Princeton University. His work focuses on ethics, Buddhist philosophy, and global priorities research.

Barry Maguire

Barry Maguire is Assistant Professor of Philosophy at Stanford University. Previously, he taught Politics, Philosophy, and Economics at UNC Chapel Hill, and held a Bersoff Fellowship at NYU. He works on issues at the intersection of normative theory, normative ethics, political ethics, and the ethics of economics.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Public Policy, Consequentialism, the Environment, and Nonhuman Animals

Mark Budolfson and Dean Spears

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.26

Abstract and Keywords

The focus of this chapter is public policy and consequentialism, especially issues that arise in connection with the environment—that is, the natural world, including nonhuman animals. We integrate some of the existing literature on environmental economics, welfare economics, and policy with the literature on environmental values and philosophy. The emphasis on environmental policy is motivated by the fact that it is arguably the most philosophically interesting and challenging application of consequentialism to policy, as it includes all the challenges of valuing the distribution of human wealth and power, and it has the further challenge of putting these consequences on the same scale as consequences for human health, nonhuman animals, and nature. We suggest that standard methods of (economic) policy analysis provide a good approximation of correct welfarist analysis, except that they must be supplemented with methods for valuing animal well-being and tradeoffs with human well-being. We then provide the needed methods.

Keywords: consequentialism, utilitarianism, policy, regulation, environment, animals, pollution, welfare, well-being, axiology

1. Consequentialism and the Environment

CONSEQUENTIALISM evaluates alternative possible courses of action (“policies” or “options” or “choices,” for short), estimates their comparative desirability, and is a leading framework for answering questions about what choices are better and worse. The focus of this chapter is on distinctive issues for consequentialism and public policy, especially those that arise in connection with the environment—that is, the natural world, and especially nonhuman animals.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The application of consequentialism to real-world decisions has three main structural components, where the latter two differ in substance between different substantive consequentialist views. These three structural components are:

(1) *impact assessment* (i.e., how different options are likely to lead to different consequences),
(2) *axiology* (i.e., how to aggregate the consequences within each possible outcome into an aggregate evaluation of the comparative goodness and badness of each outcome), and
(3) *decision theory* (i.e., how to rank options based on the probability of different outcomes conditional on their choice together with the comparative goodness and badness of those outcomes).

(p. 593) As an illustration, these factors are recognizable in familiar consequentialist debates about the ethics of individual environmentalist choices such as whether one must avoid eating meat, whether one must reduce one's carbon footprint, and the like. In each case, consequentialist analysis turns on (1) impacts (e.g., whether it makes any difference to animal suffering whether a single person buys meat or not) (Singer 1980; Norcross 2004; Kagan 2011; Budolfson 2018; Nefsky 2019), (2) how to value outcomes based on impacts (e.g., whether impacts that reduce billions of people's lives by an imperceptible amount of time could add up to something as bad as killing a normal adult in the prime of life) (Kagan 2011; Nefsky 2011; Nolt 2011; Broome 2012; Budolfson 2012), and (3) how these factors should determine what you must do (e.g., whether you are required to choose the option that maximally benefits society in expectation, or whether it is permissible to do much less as long as it is "good enough") (Singer 1972; Portmore 2011).

Much has been written from a philosophical point of view on consequentialism and these individual-level choices in response to environmental challenges. Partly for this reason, such individual-level choices will not be our focus here. In addition, we set aside individual-level issues here because they are often driven by factors that are not essentially about the environment, but by more general issues about consequentialism and demandingness, or consequentialism and collective action, and so on.

Less has been written from a philosophical point of view about consequentialism and how society should make public policy choices, especially about the environment and nonhuman animals. This will be our focus in what follows, where we endeavor to integrate some of the existing literature on environmental economics, welfare economics, and policy with the existing literature on environmental values and philosophy.

2. Standard Policy Analysis

In applying consequentialism to real-world public policy decisions, the leading method is what might be called *standard policy analysis* (SPA). As an instance of applied consequentialism, SPA combines (1) an impact assessment that models the empirical dynamics that determine outcomes as a function of policy choices (which generates an assignment of

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

probabilities to outcomes conditional on policy choices), (2) an axiology, where SPA assigns value to outcomes based on how the impacts within each outcome are valued by humans, and (3) a decision theory that evaluates the choice-worthiness of policy options as a function of (1) and (2), where SPA assumes a decision theory that has a familiar expectation or maximize expected value form, in which policy options are valued based on the sum of the possible outcomes of each policy option weighted by their probability conditional on the choice of that policy.[1] In many cases, a more partial analysis is performed (p. 594) using methods of SPA, such as a cost-benefit analysis, which might for example evaluate whether the benefits exceed the costs of a particular kind of pollution control in a number of scenarios that model various levels of stringency of control; although this is only a partial analysis that ranks the comparative choice-worthiness of a small number of policy options, the methods of this kind of cost-benefit analysis are typically those of SPA (Drummond et al. 2005).

SPA in this sense is the most influential methodology that informs public policy (Sunstein 2014; Adler and Fleurbaey 2016; Adler 2019). SPA is often (and increasingly) used in other contexts beyond public policy, such as by nongovernmental organizations (NGOs), individuals, and foundations deciding what initiatives to fund (GiveWell.org 2019), and in any other context where a decision must be made about a complex problem that can be modeled, and where valuation metrics can be designed to represent better and worse possible outcomes. SPA is especially pervasive in evaluations of policies that will have widespread socioeconomic consequences, and in the environmental domain. This is not to say that policies enacted by policymakers generally conform to SPA, as instead other values and political objectives often carry the day (Sunstein 2018). Rather, the point is merely that SPA is the most influential analytical input into actual policy analysis, and the use of SPA is widely judged to be normatively correct by widely cited scholars on public policy, even if actual policymakers rarely conform to its recommendations (Sunstein 2014; 2018).

In what follows we discuss the main substantive assumptions of SPA in more detail, highlight some common objections to SPA from philosophers and others, and explain the resources that SPA has for replying to those objections, and their limits. In later sections we discuss the prospects for improving SPA.

2.1. Impact Assessment in Standard Policy Analysis

As may already be clear, SPA factors into empirical and evaluative parts. The empirical part models the dynamics that determine outcomes as a function of policy choices, and it is generally based on work from social, health, and/or natural or other empirical sciences. For example, in the case of income tax policy, the dynamics might be taken from economic studies. In a multidisciplinary context such as air pollution policy, the dynamics might be taken from both atmospheric and public health science (namely the benefit side of the equation, based on the science of population-level impacts of different levels of exposure to air pollutants) and economics (namely the cost side of the equation, based on energy economics models of the cost of different levels of pollutant reductions via different policy instruments). In a maximally large-scale problem like climate change, an “integrated

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

assessment model" might have to be developed by teams (p. 595) of scientists from a wide variety of disciplines to model the coupled complex systems involved and their associated impacts along a multitude of diverse dimensions for different individuals at different locations in space and time, at each point estimating impacts conditional on policy choices for different sectors that drive human wealth, health, migration, and demography, and the well-being and population dynamics of flora and fauna, impacts on ecosystems, and so on. This is what is actually done in the case of climate change, and increasingly other global environmental challenges such as ecosystem preservation and the like (IPCC 2014; IPBES 2019).

We set aside the details of impact assessment given its empirical nature, but before doing so it is worth mentioning some of its limitations. The first is simply empirical uncertainty, which increases as estimates extend into the more distant future. Less obvious but important recurring problems are also a frequent inability to anticipate important negative unintended side effects of policies, and a frequent inability to estimate the capacity for innovation or social coordination to endogenously improve outcomes in response to societal challenges (Ostrom 1990; Connelly 2008; Lam 2011; Deaton 2013, chap. 7).

2.2. Axiology in Standard Policy Analysis: Anthropocentric Valuation Based on Human Preferences

Given possible outcomes modeled by an impact assessment, SPA assigns value to those outcomes based on estimates of the value of the impacts within each outcome to humans, along with a *social welfare function* that aggregates the value of all of these impacts to different humans into a single aggregate societal value of the overall outcome; together, this is the axiology of SPA. One key feature of SPA is that this axiology is *anthropocentric*, in the sense that it bases its valuation on methods of estimating how the relevant impacts would be valued *by humans*, typically measured in terms of impacts on overall gross domestic product (GDP) or societal wealth, or on as a function of the willingness to pay by individuals to achieve or avoid them.

In making these anthropocentric assumptions, SPA mirrors the normative assumption of mainstream economics about fundamental value, namely that what fundamentally makes for better or worse outcomes is the extent to which those outcomes are preferred or dispreferred by humans. It is important to see that this is indeed a *normative* assumption (since this assumption is used to conclude that some outcomes are *better* than others and thus *should* be chosen by policymakers), and that this assumption is not "neutral" (contrary to what many economists claim) since for example it implies that we should ignore the well-being of nonhuman animals in a way akin to how colonialists ignore the well-being of indigenous people except insofar as what was good for those people aligned with colonial interests.

More generally and beyond the implications for animals, many normative theorists find the axiology assumed by mainstream economics dubious, on the grounds that it (p. 596) ignores the possibility that preference satisfaction might not be the only determinant of

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

individual well-being, and because it ignores other possible determinants of better and worse outcomes, including considerations of justice (Hausman et al. 2018). Later sections will add detail to these critiques, evaluate their merit, and identify alternatives to SPA. For now, the next order of business is to better understand the assumptions of SPA so that its resources to reply to these objections can be made clearer.

2.2.1. Valuation of Impacts in Standard Policy Analysis

Because environmental challenges like air pollution and climate change can have impacts on almost the entire range of things that might be considered fundamentally valuable, it is useful to provide some framework for enumerating these valuable things, and to examine the resources that SPA has or doesn't have to properly evaluate these impacts. Here is a brief and nonexhaustive list:

- consequences for individual humans along the dimensions of:
  - wealth
  - health
  - happiness
  - freedom
  - cultural and aesthetic values
- consequences for individual nonhuman:
  - sentient animals (both wild and farmed)
  - nonsentient living things such as crops, trees, and so on
  - nonliving things such as mountains
- systemic consequences:
  - distribution of wealth, and other human distributional consequences
  - distribution of ecosystems of various compositions, and other nonhuman distributional consequences

This is not meant to be an exhaustive list, but merely indicative of what impacts might be seen as valuable given a substantive philosophical view—this list may be helpful as one considers whether SPA or some alternative approach to policy analysis can properly value all such things.

A common caricature of SPA is that it ignores the value of everything on the list except for the first thing (wealth), on the grounds that SPA cares only about things that have monetary value in the marketplace. It is important to see that this is a confused criticism, as SPA (especially in the environmental domain) often aims at valuing all of the things listed, and does so via principled and well-developed methods of valuation based on human preferences, where the essence of the project is to derive the value of things that do not have monetary values in the marketplace from other measures of human preferences.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

At the same time, it is also true that some applications of SPA are indeed overly simplistic and do not adequately accomplish this goal. In what follows, we aim to give a fair description of how SPA can be used to value things on the aforementioned list (p. 597) of things in a principled way based on human preferences, as well as a clear view of how some applications of SPA may fail to achieve this aspiration.

To begin to see why valuing things that have no monetary value in the marketplace is part of the essence of the project of SPA in the environmental realm, note that one general rationale for policy that is endorsed by proponents of SPA is what might be called the *market failure rationale for policy*, namely, that prices in the marketplace sometimes systematically fail to account for some of the costs and benefits to other humans of the transactions that give rise to those prices, and under these conditions unregulated market transactions will not generally lead to the best outcomes for society, and so regulatory intervention is justified. Most on point for our purposes are cases where *negative externalities* of transactions exist, in the sense that some individuals are harmed by a transaction because they are not party to the transaction and thus the price at which the exchange happens does not reflect the strength of their preferences. For example, in the 1950s air pollution emissions were largely unregulated in the United Kingdom and the United States, and as a result, when a factory owner produced goods and sold them in the marketplace and created emissions in the process, the preferences for air quality of all the people who were harmed by those emissions were generally ignored in determining the market prices at which the goods were bought and sold. And if everyone's preferences for air quality had been taken into account (i.e., if the producer had to pay everyone harmed by his or her pollution in the same way consumers had to pay the producer for the goods he or she produced, namely, to the point at which individuals were happy to accept the air pollution in exchange for that payment), then some factories would have had to close and the air would have thus ended up cleaner, and the value of the air quality benefits (measured in the aggregate willingness to pay for them across society) would have been more than enough to compensate for the lost production. In this way, the lack of regulation of air pollution in the 1950s created a situation where free markets led to outcomes that were worse than what they could have been if everyone had to pay the true social cost of their pollution. If prices had instead internalized all of the true costs in the way just described, then many people could have been made better off without anyone being made worse off, a *pareto improvement*, generally taken by economists to be an uncontroversial example of a better outcome.

One of the contributions of welfare economics is the proof that under conditions of perfect competition including no externalities, the end result of free exchanges in society would be an outcome that is *pareto optimal* in the sense that no pareto improvements to that outcome would be possible; at the same time, when negative externalities exist or other features of imperfect competition exist such as monopoly power, then there is a clear reason to expect *market failure* in which free exchange would result in a suboptimal outcome in the sense that a pareto improvement would be possible (Kolstad 2010; conceptually this is based on the *First Fundamental Theorem of Welfare Economics*). In the face of market failure, a rationale emerges to use regulatory policy to improve outcomes.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

It is important to see that the nature of this market failure rationale is pro tanto and may be defeated in many cases, because policy should not always be expected to improve outcomes, as government policies often generate even worse outcomes, (p. 598) given that government is imperfect in predictable ways just as unregulated markets are imperfect in predictable ways (Budolfson 2017).

A conceptually instructive policy instrument that is often considered by SPA as a means to correct market failures involving environmental pollution is a *Pigovian tax*, which is price charged to those who impose pollution (or more precisely, negative externalities on others) based on a calculation of the aggregate cost to society (the *social cost*) of that type of pollution, measured in monetary terms by aggregate willingness to pay across society to avoid that type of pollution. Theoretically, if such a tax were levied at the marginal cost to society of an additional unit of the relevant type of pollution at the point at which that marginal cost is equal the marginal benefit to society of an additional unit of that form of pollution, then the optimal level of that form of pollution would result (Kolstad 2010).

Returning now to the long list of potentially valuable impacts, the key resources of SPA are methods that aim to value all such impacts on a single scale of willingness to pay that represents aggregate human preferences to achieve or avoid those impacts, even when those impacts do not have a market price. When done well and comprehensively, SPA thus promises to value all valuable impacts in a principled way based on human preferences. Furthermore, according to mainstream economics, this is the only normatively acceptable way of valuing impacts, since policies based on any other valuation scheme would illegitimately force outcomes onto society that do not reflect the preferences of the population, and instead involve dictators imposing their own values on the population (Nordhaus 2007, 691). We consider objections to this view further later. For now, we outline in more detail the methods SPA uses to value the impacts enumerated earlier in connection with a representative range of leading examples in environmental policy and philosophy:

- environmental pollution
- the treatment of nonhuman animals
- natural resource use and land management
- environmental justice (discussed in the next subsection)

Environmental pollution has been mentioned already, but the valuation of impacts has not been discussed at the level of detail relevant to understanding best practices applications of SPA. In the case of pollution, applying SPA requires identifying and modeling the relevant impacts of pollution and then valuing them. Some of the impacts typically modeled are mortality (e.g., additional deaths from heart attacks as a function of increased exposure particulates in the air) and morbidity (e.g., additional burden of asthma), impacts on recreation and cultural and aesthetic values (e.g., preference for clear skies for aesthetic reasons, and a preference for cathedrals not to be defaced by acid rain), and pollution has impacts on many other things such as crop yields and the like. SPA aims to monetize each of these impacts based on human willingness to pay, and thus put them on the same scale

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

as the benefits from goods produced by the emissions-generating activities. The general aim of policy is then to reduce emissions in a way that (p. 599) reflects all of these preferences and associated willingness to pay, down to a level at which market failure no longer generates a suboptimal outcome.

In the case of farm animals, SPA also has the resources to provide a market failure rationale for animal welfare improvements, as many economists argue that humans are sufficiently willing to pay for better animal welfare to more than offset the cost of some animal welfare improvements (Cowen 2006; Norwood and Lusk 2011). A different market failure rationale for animal welfare improvements is that they would more than pay for themselves by reducing the expected harm to human health from diseases, antimicrobial resistance, and the like, where these harms to human health are not reflected in the market prices of animal products; thus, policies that included targeted animal welfare improvements could yield benefits for everyone in expectation (Otte and Chilonda 2000; Jarvis and Donoso 2018). In this way, SPA has the resources to argue for substantial animal welfare improvements—namely, to a higher level of animal welfare that best satisfies the preferences of humans.

This provides a useful segue into an outline of general SPA methods for estimating willingness to pay for *nonmarket* impacts that do not correspond to market prices. Valuing these impacts is a nontrivial challenge. One dimension of this challenge is estimating the *use value* of elements of nature (human willingness to pay to use animals, trees, minerals within a mountain), and another dimension is estimating the *nonuse value* (willingness to pay for such things to remain unused). Use value includes willingness to pay for direct use (e.g., to buy and eat animals, convert trees into housing materials), as well as indirect use, including *ecosystem services* such as the value of pollinators in human agriculture, the value of aquatic mollusks in cleaning water for human use, the value of wildlife to human recreation, and so on. Nonuse value also includes willingness to pay for the continued mere existence of elements of nature without use by humans (*existence value*, which is especially important for preservationist valuation of wildlife, biodiversity, wilderness areas), as well as the *option value* of keeping elements of nature around for potential future human use in ways that will turn out to be valuable, but of which we may be currently ignorant (Arrow and Fisher 1974).

Substantive methods are needed to estimate willingness to pay for these things, as their values are often not readily reflected in market prices.[2] *Contingent valuation methods* are generally based on surveys that elicit self-reported willingness to pay to avoid or bring about particular outcomes. However, there are a number of objections to this method, perhaps the most important of which is the worry that it leads to biased and inflated estimates of willingness to pay, because people do not have to back their answers with real investments (Hsiung and Sunstein 2007). This points the way toward the main alternative method, based on *revealed preference methods* for measuring valuations implicit in actual choices, beyond what can be immediately read off from market prices. Revealed preference studies are especially foundational in the valuation of mortality, as (p. 600) willingness to pay to avoid increased occupational risk of death and the like are often used to es-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

timate the value of excess deaths or life years lost. These estimates of willingness to pay for human health are often the dominant factor in estimates of the social cost of pollution and other cases where environmental quality has a clear impact on human health (Sunstein 2014). These methods can also be used to infer willingness to pay for amenities like open space, parks, and the like when those amenities can be seen as partly determining the price of things like housing that do have clear market prices; in such cases, inferential methods can be used to extract an estimate of the contribution of the amenity to the comparative prices of, for example, houses.

These methods are the most widely used methods for estimating the value of nonmarket impacts (Kolstad 2010).

Policy analysis also often requires methods of estimating willingness to pay for resources in the future, and a calculation of the present value of those future benefits, as ignoring these long-run benefits in an unregulated market could create a tragedy of the commons in which resources are used unsustainably with an eye toward only short-term profits, leading to a worse outcome for society than if they were managed in accord with an analysis that accounted for long-run value (Ostrom 1990; Sandler 2018, 113–124).

2.2.2. Social Welfare Functions in Standard Policy Analysis

Given the valuation within each possible outcome of impacts to individuals by SPA, the next step is to aggregate those into an overall valuation of each outcome, so that the goodness and badness of the different possible outcomes of policy choices can be compared. There is no single agreed-upon formula—*social welfare function (SWF)*—for this aggregation. Instead, there are a number of different SWFs that are sometimes used in SPA, which we will describe and contrast. All these SWFs are motivated by the conceptual idea of *individualist anthropocentric welfarism*, in the sense that they share the basic conceptual idea that the goodness of an overall outcome is a principled function of the well-being of the individual humans within that outcome, and that individuals have more well-being the more they consume, where this notion of an individual's *consumption* is taken to include all the goods and services, leisure time, health, and everything else that the individual values, as measured by the individual's willingness to pay.

An important distinction between different SWFs used in SPA is whether they merely focus on a societal level economic sum (e.g., GDP plus the net monetized value of all of the nonmarket impacts described in the previous section), or whether instead they represent differences between individuals, such as different levels of individual consumption, differences in race, gender, age, location, and so on, and accordingly estimate different well-being consequences when the same impacts affect different individuals, and then aggregate those heterogeneous well-being consequences. It is common to use the former method, although the latter is more precise from a normative point of view. Some advantages of the former method are that it is simple and avoids the need for an assumption about how to make interpersonal comparisons of well-being levels. However, if one accepts that there is diminishing marginal utility of consumption in the (p. 601) case of any given individual (as every economist does), one would presumably think that at the popu-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

lation level a similar diminishing marginal utility of consumption would arise—that is, one would assume that at the population level giving an additional dollar of consumption to the poor would tend to increase well-being more than giving that additional dollar to the rich. A widely used SWF in SPA that captures this thought is the following *isoelastic utility function*:

(1)

$$W^{TU} = \sum_{i \in humans} \frac{(c_i)^{1-\theta}}{1-\theta}$$

For our purposes, the important feature of the SWF in Equation 1 is that there is diminishing marginal utility of consumption, which allows a policy analysis to model the idea that there is greater well-being generated by a dollar worth of increased consumption for the poor versus the rich. Empirical studies of reported happiness and income provide some evidence for kind of approach as well (Kahneman and Deaton 2010). In addition, this approach avoids the normatively objectionable implication that impacts to the poor are less important than impacts to the rich simply because the poor have lower willingness to pay, which is an implication of the first approach that simply sums monetized values of impacts.

2.2.3. Enlightened Standard Policy Analysis

With all of this in mind, many advocates of SPA believe that best practices should use a SWF analogous to Equation 1 in which individual well-being is estimated as a concave transformation of consumption, which can then be aggregated at the societal level in any number of ways that can include principles of distributive justice, while avoiding the implication that the poor are worth less simply because of their lower willingness to pay for health, life, and other goods (Adler and Fleurbaey 2016; Adler 2019). Although Equation 1 is a total utilitarian SWF, alternative SWFs exist and have been advocated as best practices within SPA to represent prioritarian, egalitarian, Rawlsian maxi-min, and other methods of aggregation, as well as the range of nontotalist welfarist population ethics (Budolfson and Spears 2018). Because the choice of SWF involves the choice of a population ethics as well, this can have implications for policy in some cases (Scovronick et al. 2017), but not in a way that need be any more dramatic than the familiar way that the choice of the shape of the transformation between consumption and well-being has implications for policy (Budolfson and Spears forthcoming). In addition, although it is commonly believed that totalist SWFs have a special liability to the *repugnant conclusion* (when an axiology implies that an enormous number of barely worth living lives can be better than a smaller number of good lives), recent work has suggested that any SWF that endorses tradeoffs between the well-being of individuals will have such implications, and so the repugnant conclusion arguably does not tell for or against any welfarist SWF (Budolfson and Spears forthcoming; 2018). (See the discussion of insect valuation and the repugnant conclusion later.)

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The upshot is that SPA need not ignore considerations of distributive justice and future generations, and indeed it has the resources to integrate them into policy analysis—we (p. 602) might call such an approach *Enlightened SPA*.[3] At the same time, it should be emphasized that in actual practice it is more common for SPA to use a normatively inferior SWF that aims merely to maximize the sum of GDP, plus some monetized nonmarket impacts—and outside the environmental realm, it is very common for applications of SPA to focus only on maximizing GDP, without any accounting for nonmarket impacts, let alone the socioeconomic distribution of impacts (Stiglitz et al. 2009).

In connection with environmental justice, distinctive issues arise about the distribution of environmental impacts. In particular, a familiar concern is that racial minorities are exposed to more pollution and other environmental inequalities, such as having fewer recreation opportunities and more waste dumps located near their homes (Shrader-Frechette 2002). In response, some would argue that this may be merely a consequence of economic inequality, as the price of homes will be lower in environmentally undesirable locations, and so it is an inevitable result of free choice by the poor to do as best as they can within the marketplace to live disproportionately in such locations (Banzhaf 2009). However, recent work has provided some evidence that this explanation may be inadequate, as in the United States, minorities are disproportionately exposed to air pollution even after controlling for income differences (Mikati et al. 2018). An Enlightened SPA SWF that displays aversion to inequality in health-related impacts could be used to capture the importance of alleviating such injustice.

A further dimension of complexity is that many environmental policy challenges such as climate change require policy analysis over a very long time horizon, with benefits distant in the future from the cost of investments to achieve those benefits. In such a case, an SWF must incorporate key assumptions about how to calculate the present value of the future costs and benefits for society, often referred to as assumptions about *discounting*. One dimension of discounting is the parameter θ, which parameterizes the diminishing marginal utility of consumption; another dimension of discounting not represented here is the rate of pure time preference, which determines how much less weight should be given to well-being consequences in the future simply because they are in the future. Some economists advocate particular methods of estimating these parameters based on what is allegedly a revealed preference methodology, which has been contested at length by economists, philosophers, and others (Nordhaus 2007; Fleurbaey et al. 2019).

In environmental discourse, *sustainability* is frequently cited as a goal, in roughly the sense of meeting the needs of the present without compromising the ability of future generations to meet their own needs (Brundtland 1987; Solow 1991). Endorsing sustainability as the *ultimate* goal of policy would imply that policy should aim not to maximize (discounted) expected value into the future (as in standard SPA), but rather to do "good enough" for future generations. SPA can be modified via an alternative SWF to encode such a sustainability objective (Fleurbaey 2015).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

This section has shown that SPA has resources to account for many values at stake in challenges familiar from environmental policy and philosophy. More generally, this (p. 603) section has provided a conceptual overview of SPA, aimed at providing a suitable background for philosophical engagement. For more details, see leading textbooks and other resources on environmental economics, social choice and welfare, and health economics (Drummond et al., 2005; Kolstad 2010; Adler and Fleurbaey 2016; Adler 2019).

3. Critiques of Standard Policy Analysis

A consequentialist view typically grounds all value in consequences for normatively relevant individuals, where the relevant individuals and the value of the consequences is explained by a theory of what is fundamentally valuable. Common views about what is fundamentally valuable include hedonism, preference satisfaction, objective list views of well-being (including possession of particular capabilities), and (more common in environmental philosophy) biocentric views on which being fulfilled as a living organism is fundamentally valuable. These views share the structural assumption that there is some set of individuals that fundamentally matters (e.g., on a hedonist view, those that can experience pleasure and pain), whereas other individuals do not fundamentally matter. In the environmental realm, holism is also a common view, on which ecosystems and other holistic entities have fundamental value (Hiller 2014; Sandler 2018). Consequentialist views sometimes also assign fundamental value to inequality and other properties of *distributions* of good and bad consequences to individuals as well (Adler and Fleurbaey 2016; Adler 2019).

3.1. Is Standard Policy Analysis Inadequate If the Preference-Satisfaction View Is Not Ultimately Correct?

From the discussion in the previous section, it is clear that SPA has impressive resources to value outcomes in terms of human preference satisfaction. At the same time, given its focus on human preference satisfaction, consequentialist philosophers are often quick to argue that SPA is inappropriate for policy analysis, on the grounds that the correct consequentialist view is not the anthropocentric version of the preference satisfaction view.

However, this argument is too quick. The problem is that even if we assume for the sake of argument that the correct fundamental view is not the preference satisfaction view, it could still be true that SPA is our best option for policy analysis and provides very reliable estimates. As an example to illustrate, human preference satisfaction as measured by SPA might be so closely correlated with an anthropocentric objective list view of well-being that there could be little extensional difference between what SPA implies and what an analysis using that objective list theory would imply—and furthermore, while we actually have methods for doing analysis using SPA, in contrast (p. 604) we arguably do not have similarly adequate methods for using the objective list theory directly. If that were all true, then from the perspective of such an objective list theory there would be no objection to the use of SPA for policy analysis, and in fact it would be a mistake not to use SPA.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

As one possible realistic example of this type, one might imagine an objective list theory of well-being in connection with policy challenges such as how best to promote the UN Sustainable Development Goals, which include goals that closely correspond to items on many objective list theories. Because there will be synergies and tradeoffs between these goals that have to be analyzed in a maximally complex system of coupled human and natural systems (Nilsson et al. 2016), and because there is arguably no better theory of how to make tradeoffs between goals on the objective list beyond what can be implemented using Enlightened SPA, policy analysis that uses Enlightened SPA might provide the best feasible estimates of policy questions such as how to best invest scarce resources to promote these goals, even assuming the truth of an objective list theory. Upon reflection, all of this should come as no surprise, given the familiar point that the best models for practical purposes sometimes rely on false simplifying assumptions about fundamental facts.

The upshot is that whether or not SPA is adequate for policy analysis in a particular context depends on whether a SPA analysis can, in that context, provide a sufficiently reliable approximation of the correct consequentialist analysis—and in many contexts, it seems that we might indeed expect Enlightened SPA to provide a sufficiently reliable approximation.

3.2. A Fruitful Approach to Improving Policy Analysis

With the preceding in mind, a more fruitful approach to improving policy analysis is not to dismiss SPA out of hand on the grounds that the correct fundamental values are not explicitly represented, but rather to focus on the question of whether SPA is good enough in particular contexts, or whether instead SPA can feasibly be improved to approximate better what ultimately matters. If we believe SPA is inferior to some feasible alternative, our critique should then aim to make such an alternative precise and readily implementable in actual policy analysis. In particular, we should characterize precisely how the relevant tradeoffs should be made by the axiology, and what decision theory should be used. If our critique lacks this level of precision, then it fails to offer any helpful recommendation for how to improve policy analysis.

As a simple example to illustrate, a theorist could argue that health should be valued more highly relative to other goods than it is in SPA, on the basis of an argument that it is more important to what ultimately matters than is indicated by human preferences alone. If this is to have relevance to policy analysis, a precise account should be offered of how health should be weighed and traded off with other goods on this view (or what range of assumptions should be used to test the sensitivity of policy recommendations). Or, as another example, a theorist could argue that ecosystem health has great fundamental value. If this is to have relevance to policy analysis, a precise formula for measuring (p. 605) ecosystem health should be specified in a way that is implementable in policy analysis. As these two examples indicate, different consequentialist views may generate different and sometime incompatible recommendations for how policy analysis should be modified.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

In some cases, there might be broad consensus among leading normative theories as to how policy analysis should be improved. As one example from the previous section, there is presumably broad consensus among leading normative theories that Enlightened SPA is an improvement over a version of SPA that focuses merely on maximizing GDP. In the next section, we provide another example of an improvement that is possible over even best practices Enlightened SPA, related to animal welfare. This also provides a worked example of how normative theorists might aim to have positive impact on policy analysis by working directly on the details of such improvements and providing precise methods that can be readily implemented in actual policy analysis.

3.3. A Fundamental Problem with Standard Policy Analysis: Anthropocentrism

There is widespread consensus among normative theorists that an important problem with even the best anthropocentric methodology is that animals and other aspects of nature within such a methodology are always valued merely in terms of their value to *humans* (Ng 1995; Jamieson 2008; O'Neill et al. 2008; Gruen 2011; Sarkar 2012; Hiller et al. 2014; Palmer et al. 2014; McShane 2018; Sandler 2018; Schmidtz and Shahar 2018). In other words, SPA valuation is always in terms of the ultimate value of outcomes to humans only, and thus assigns no fundamental value to the well-being of animals.[4] For example, on even the best anthropocentric approach, the deaths of billions of birds due to climate change would have disvalue only insofar as the deaths of those birds have disvalue to humans. Most normative theorists would object that this way of valuing animal lives is fundamentally incorrect because it ignores the value of the birds' own well-being irrespective of its contribution to human well-being, as scientists and theorists broadly agree that animals, like humans, experience different levels of well-being depending on decisions made by others, and there is no normatively principled argument that the well-being of animals should be ignored while that of humans should not (Singer 1975; Kagan 2019). Further, the assumption of anthropocentrism is dubious even from within the logic of mainstream economics, since sophisticated animals have preferences over outcomes and there is nothing within economics that explains why the preferences of this subset of individuals should be ignored, just as there was never an economic logic to ignoring the preferences of people in an earlier time based on their race, gender, or other (p. 606) factors. As a result, normative theorists generally agree that the well-being of animals must be included in any full accounting of the well-being consequences of decisions.

3.3.1. The Challenge of Interspecies Comparisons

Animal welfare is almost never included in policy analysis, partly due to methodological prejudice, but increasingly also because we do not currently have good methods for quantifying animal well-being consequences and putting them on the same scale as quantified human well-being consequences. We might call this "the challenge of interspecies comparisons."

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Recent work by Kevin Wong has highlighted the most difficult problem that needs to be solved in connection with interspecies comparisons, which is how to estimate the *well-being capacity* (well-being potential) of members of a nonhuman species relative to the well-being capacity of humans (Wong 2016). If we knew how to make those interspecies comparisons of well-being capacity, then we could integrate animal welfare consequences into existing methods of decision analysis, by deriving empirically based estimates of animal welfare consequences on the same scale as human consequences that typically underpin welfarist decision-making analyses.

For example, suppose an additional degree of climate change will cause us to lose 1 million life years of a particular species of bird, and we want to value this on the same scale as losses to human life from an additional degree of warming that are already modeled and valued based on an assumption about the value of one human life year. If we had a good estimate of the well-being capacity of that species of bird relative to a human, we could then multiply that estimate by the purely empirical impact estimate of 1 million life years lost to get an estimate of the amount of well-being lost by that bird species on the same value scale as the existing estimate of human well-being loss, assuming that one degree of additional climate change does not change the quality of life of those birds. And if one additional degree of warming does diminish the quality of life of the remaining birds of that species, we can simply the multiply the number of remaining bird species life years by a further *quality of life adjustment* term that is itself an empirical impact estimate from zoological experts and the like. (We can also use such a term to take into account any antecedent diminishment in the well-being experienced by all of the birds including those that would die before the warming.)

This line of thought leads to the following equation for the average well-being experienced by a member of a species s (which we symbolize as $\overline{u}_s$) as a function of the average well-being capacity per unit of time of members of s relative to humans ($\overline{\pi}_s$), multiplied by the average duration of a life of a member of s ($\overline{\delta}_s$), multiplied by a quality of life adjustment term ($\overline{f}_s$):

(2)

$$\overline{u}_s = \overline{\pi}_s * \overline{\delta}_s * \overline{f}_s$$

The key point here is to highlight the term $\overline{\pi}_s$ as the key unknown term, where the unsolved problem of how to estimate $\overline{\pi}_s$ is the essence of the challenge of interspecies comparisons. (The other terms $\overline{\delta}_s$ and $\overline{f}_s$ are susceptible to existing empirical methods, (p. 607) where the term f can be seen as the focus of existing animal welfare science—see, for example, Fraser 2008; Appleby et al. 2011, and Browning 2019.[5])

3.3.2. A Method for Quantifying Animal Welfare and Making Interspecies Comparisons

In this section we propose a method of making interspecies comparisons that has some analogy to the method used in Equation 1 of taking consumption as a proxy for human

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

well-being: the proposal is to make interspecies comparisons based on a proxy that is imperfect but delivers estimates as good as we can expect in practice. To do this we first identify a proxy, call it n, to use as the basis for estimating well-being potentials across species, analogous to the use of consumption (c) as the basis for estimating well-being across humans. As an overly simplistic illustration of this idea, n might be the number of neurons in the brain of members of a species. Data on number of neurons are readily available, and they may be a good proxy in some select contexts, such as an enormous global analysis involving billions of individuals where different species are crudely lumped together in small number of bins such as "mammals" and "insects." When greater accuracy is required for specific species or individuals, n can be set equal to a more complex metric based on expert analysis of empirical properties that are best correlated with different levels of well-being (which might differ according to different substantive theories of well-being)—for example, the number of neocortex-like neurons, cortisol levels, sociality, or other leading factors identified by the scientific community and philosophers as most closely correlated with well-being capacity (Fraser 2008; Appleby et al. 2011; Dawkins 2012; Shriver 2014; Barron and Klein 2016; Klein and Barron 2016; Olkowicz et al. 2016; Herculano-Houzel 2017; Tye 2017).

Abstracting from those details, which are not essential to the core challenge of how to make interspecies comparisons, the first step of the proposal is to parameterize such an empirical proxy n, perhaps with an exponential weight ψ, into estimates of comparative well-being capacity for different species. The second step is to multiply this estimate of well-being capacity by a descriptive measure of the degree to which this potential is actually realized, and multiply by the f quality of life and δ duration terms, to yield the desired well-being estimates. For example:

(3)

$$W^{TU} \approx \sum_{is} n_{is}^{\psi} f_{is} \delta_{is}$$

(In ordinary language: the total sum of well-being is approximately equal to the sum over all individuals across species of that individual's empirical basis for well-being capacity raised to the normative exponent [which determines the relationship between the empirical proxy and well-being capacity] multiplied by the quality of life adjustment, multiplied by the duration of that individual's life.)

(p. 608) In practice, it would often be more feasible without important loss of accuracy to use species-level averages (where averages are denoted by a bar over the letter) as the proxy for well-being potential, which can then be multiplied by the species population P_s:

(4)

$$W^{TU} \approx \sum_{s} P_s \overline{n}_s^{\psi} \overline{f}_s \overline{\delta}_s$$

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Equations 3 and 4 summarize the proposed method for making interspecies comparisons. They require an empirical proxy for n (e.g., number of neurons, or a more complex empirically based metric), values for the normative parameter such as ψ that are grounded in normative and empirical considerations (on analogy with how values for θ in Equation 1 is grounded in normative and empirical considerations), and empirically determined values for f and δ.

Note that Equation 4 provides a practical method of estimating the value of the following more theoretically obvious equation that multiplies the population P_s of each species by the average well-being $\overline{u}_s$ of members of that species:

(5)

$$W^{TU} = \sum_s P_s \overline{u}_s$$

The problem of interspecies comparisons means that we cannot directly use Equation 5 prior to a method of making interspecies comparisons such as that developed earlier, as using Equation 5 directly would require knowing the value of $\overline{u}_s$ for each species, which would require knowing the answer to the question of how to make interspecies comparisons. Instead, we must first pioneer a method for making those comparisons, such as provided by Equations 3 and 4: namely, to take $\overline{u}_s \approx \overline{n}_s^{\psi} \overline{f}_s \overline{\delta}_s$.

Note that if we were to translate Equation 2 into a total utilitarian axiology, this would yield:

(6)

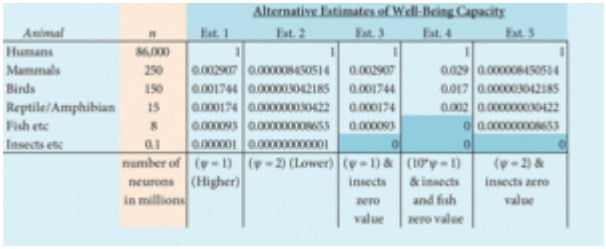
$$W^{TU} \approx \sum_s P_s \overline{\pi}_s \overline{f}_s \overline{\delta}_s$$

When we substitute the term $\overline{n}_s^{\psi}$ for $\overline{\pi}_s$ in Equation 6, the result is Equation 4. A term like $\overline{n}_s^{\psi}$ can similarly be incorporated into other population axiologies in a straightforward way, but for ease of exposition we focus only on totalism in this chapter. (See Budolfson and Spears 2018, forthcoming, and 2019b, for further discussion of population ethics, including in connection with animal welfare.)

3.3.3. Estimates of Optimal Tradeoff Rates between Humans and Animals, and the Repugnant Conclusion

Figure 31.1 illustrates how a sensitivity test could be incorporated into policy analyses based on different principled ways of using the parameter ψ to estimate potential well-being of a species s as a function of the average number of neurons *n* in a member of that species:

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).



		Alternative Estimates of Well-Being Capacity				
Animal	*n*	Est. 1	Est. 2	Est. 3	Est. 4	Est. 5
Humans	86,000	1	1	1	1	1
Mammals	250	0.002907	0.000008450514	0.002907	0.029	0.000008450514
Birds	150	0.001744	0.000003042185	0.001744	0.017	0.000003042185
Reptile/Amphibian	15	0.000174	0.000000304220	0.000174	0.002	0.000000304220
Fish etc	8	0.000093	0.000000008653	0.000093	0	0.000000008653
Insects etc	0.1	0.000001	0.000000000001	0	0	0
	number of neurons in millions	($\psi = 1$) (Higher)	($\psi = 2$) (Lower)	($\psi = 1$) & insects zero value	($10^*\psi = 1$) & insects and fish zero value	($\psi = 2$) & insects zero value

Figure 31.1. Five alternative estimates of the well-being potential of animal life years of different species based on the number of neurons in an average member of the species. Each estimate is expressed in terms of the well-being capacity of one human life year, and thus each estimate divides by the estimated well-being capacity of one human life year, $\overline{n}_h^{\psi}$. Estimate $1 = \frac{\overline{n}_s^{\psi}}{\overline{n}_h^{\psi}}$ with ψ set equal to 1 (a higher estimate of the capacity of animals), whereas estimate $2 = \frac{\overline{n}_s^{\psi}}{\overline{n}_h^{\psi}}$ with ψ set equal to 2 (a lower estimate of the capacity of animals). Estimates 3 and 5 both stipulate that insects have zero capacity for well-being (with the rationale that they fall below some critical threshold), but otherwise use estimates 1 and 2, respectively. Estimate 4 assumes both insects and fish have zero capacity but adds a much higher estimate of the capacity of other animals by multiplying the estimate 1 fraction by 10 for mammals, birds, reptiles, and amphibians.

Each estimate can be used to put human life years (which can be estimated via familiar proxies such as Equations 2 or 3) on the same scale as the life years of animals of (p. 609) different species, and each estimate does so in a principled way that is empirically grounded. For example, assuming number of neurons as a basis for well-being estimates, if ψ is set equal to 2 (a principled lower value for animals), then a human life year is worth almost 120,000 mammal life years, and almost 120,000,000 fish life years. If instead ψ is set equal to 1 (a principled higher value for animals), then a human life year is worth about 344 mammal life years, and about 10,700 fish life years. These alternative estimates appear to represent much of the range of empirically grounded and principled views over the well-being of animals of different species (Herculano-Houzel 2017), and they can avoid unintuitive implications. It may not even be desirable to attempt to choose between these estimates in policy analysis, if the goal is to take normative uncertainty into account and test the sensitivity of optimal decisions to this range of different reasonable (and empirically and theoretically principled) estimates.

Of particular note are the implications (and lack thereof) for the repugnant conclusion. One might think a priori that assigning any positive value to the lives of insects could dominate analyses in a repugnant conclusion-like way because there are many quadrillions of insects (Singer 2016; see also Tännsjö 2016 and Fischer 2016). For a real-world example, insects are estimated to benefit in numbers and in average well-being

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

from climate change, and so there is some worry that if they are assigned any positive value, that will imply that we should do nothing about climate change (Sebo forthcoming).

However, using the valuations from earlier, it can be shown that a repugnant conclusion need not be an implication of assigning positive value to insects in a principled way. (p. 610) For example, the method in the previous section provides a way of demonstrating that this need not be true for principled valuation scheme with $\psi = 2$, which assigns value to insects using the same principled function of number of neurons as it uses to assign value to other animals, and in a way that seems to capture arguably a very common view in society about how to assign comparative weight to the interests of animals versus humans. When these higher ψ values are used in climate policy analysis, insects do not dominate the calculation because they are assigned such miniscule value in a principled way by the ψ exponent (Budolfson and Spears 2019a). Still, under $\psi = 1$ (a higher valuation of animals that is uncommon in general society), it is true that insects can dominate the calculation because of their sheer numbers together with that higher valuation of their lives. So adding valuation of animal well-being in a principled way does not necessarily lead to a repugnant conclusion if insects are given positive well-being in a principled way (e.g., $\psi = 2$), but insects can indeed "repugnantly" dominate at higher valuations of "lower" animals (e.g., $\psi = 1$). This is an instance of the general possibility for every approach to tradeoff-making social evaluation to yield repugnant-seeming outcomes when applied to large numbers (Budolfson and Spears 2018).

In sum, the method outlined here allows interspecies comparisons based on empirically available estimates of species population dynamics and within-species quality of life adjustment, together with empirical proxies for well-being capacity n that can be calibrated with the ψ parameter to reflect normative uncertainty about the connection between those empirical proxies and well-being capacity. Implementing these methods in policy analyses would have an important impact on estimates of how best to invest time and money by individuals (Budolfson 2015), businesses (Berkey forthcoming), and charities (including for purposes of "effective altruism") (Wong 2016; ACE 2019; OPP 2019), and similarly for estimates of optimal public policies for correcting market failures (Cowen 2006; Norwood and Lusk 2011; Jarvis and Donoso 2018), for sustainable intensification of agriculture that aims to take animal welfare into account (producing more food while reducing the overall impacts of agriculture) (Garnett et al. 2013), for climate change policy (how quickly we should be reducing greenhouse gas emissions) (Hsiung and Sunstein 2007; Budolfson and Spears 2019a), and for wilderness protection policy and other challenges related to natural resource management (Hsiung and Sunstein 2007; Sunstein 2018, chap. 6; Fischer et al. forthcoming). In all of these cases, if the well-being of animals is taken more fully into account, then decisions by individuals, governments, and others will become better on welfarist grounds.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

4. Conclusion: A Perspective on Consequentialism and Policy Analysis

Enlightened SPA is a powerful first step toward correct consequentialist policy analysis. However, it is not capable of including fundamental valuation of the well-being of (p. 611) nonhuman animals, and thus must be supplemented with methods for including their well-being, at least if the goal is to provide a philosophically defensible form of welfarist policy analysis. Many other important objections were briefly noted earlier or in some cases not discussed in detail: for example, some would argue that welfarist policy analysis should be constrained by fairly strict "side constraints" to respect basic human rights and less stringent constraints to respect the outputs of institutions necessary for wealth creation, including property rights, corrective justice, and free exchange. These constraints could themselves be justified on consequentialist grounds as optimal features of the basic structure of society. Perhaps policy analysis should also be constrained by extreme modesty about the limitations of impact assessments to include unintended consequences and the likelihood of endogenous solutions not comprehended by policy projections.

A more controversial perspective is that together, these constraints create a default toward a form of classical liberalism, where this default is often overridden in particularly clear cases such as air pollution regulation and climate change, or in the provision of basic public goods, including funding basic research on medicine and public health, and perhaps a social minimum of resources needed for a healthy life (Budolfson 2017). When the default is overridden and regulation is needed, (constrained) welfarist consequentialism is appropriate for policy analysis, implemented via advanced social welfare functions of the sort advocated here, which can capture at least most dimensions of value in most cases, including for animal welfare, valuing inequalities and other considerations of distributive justice, and thus provide the best tools for realistic policy analysis in complex societies.

A key remaining question is what the correct parameters are to use in these social welfare functions to value aversion to inequality, the comparative well-being capacity of humans and other species, and other key parameters for determining social aggregation along which there is currently normative uncertainty. These are key issues for further research. By working directly on the details of calibrating these parameters, and delivering precise proposals for other improvements that can be readily implemented in policy analysis, normative theorists can aim to have positive impact on policy analysis.

References

Adler, Matthew. 2019. *Measuring Social Welfare*, New York: Oxford University Press.

Adler, Matthew, and Fleurbaey, Marc, eds. 2016. *Oxford Handbook of Well-Being and Public Policy*. New York: Oxford University Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Appleby, Michael, Mench, Joy I., Olsson, Anna, and Galindo, Francisco, eds. 2011. *Animal Welfare*. 2nd ed. London: CABI.

ACE 2019. "Animal Charity Evaluators." **https://www.animalcharityevaluators.org/**.

Arrow, Kenneth, and Fisher, Anthony. 1974. "Environmental Preservation, Uncertainty, and Irreversibility." *The Quarterly Journal of Economics* 88, no. 2: 312–319.

Banzhaf, H. S. 2009. "The Political Economy of Environmental Justice." *Resources*, May 25.

Barron, Andrew, and Klein, Colin. 2016. "What Insects Can Tell Us about the Origins of Consciousness." *Proceedings of the National Academy of Sciences* 113, no. 18: 4900–4908.

Berkey, Brian. Forthcoming. "Prospects for an Animal-Friendly Business Ethics." In *Animals and Business Ethics*, edited by Thomas. New York: Springer.

(p. 612) Broome, John. 2012. *Climate Matters*. New York: Norton.

Browning, Heather. 2019. "If I Could Talk to the Animals: Measuring Subjective Animal Welfare." PhD diss., Australian National University.

Brundtland, Gro. 1987. *Report of the World Commission on Environment and Development: Our Common Future*. United Nations General Assembly document A/42/427.

Buchak, Lara. 2013. *Risk and Rationality*. New York: Oxford University Press.

Budolfson, Mark. 2012. "Collective Action, Climate Change, and the Ethical Significance of Futility." **http://www.budolfson.com/papers**.

Budolfson, Mark. 2015. "Consumer Ethics, Harm Footprints, and the Empirical Dimension of Food Choices." In *Philosophy Comes to Dinner*, edited by A. Chignell, T. Cuneo, and M. Halteman. New York: Routledge.

Budolfson, Mark. 2017. "Market Failure, the Tragedy of the Commons, and Default Libertarianism in Contemporary Economics and Policy." In *Oxford Handbook of Freedom*, edited by David Schmidtz and Carmen Pavel. New York: Oxford University Press.

Budolfson, Mark. 2018. "The Inefficacy Objection to Consequentialism and the Problem with the Expected Consequences Response." *Philosophical Studies* 176:1711–1724.

Budolfson, Mark, and Spears, Dean. Forthcoming. "Does the Repugnant Conclusion Have Important Implications for Axiology or for Public Policy? In *Oxford Handbook of Population Ethics*, edited by Gustaf Arrhenius, Krister Bykvist, and Tim Campbell. New York: Oxford University Press.

Budolfson, Mark, and Spears, Dean. 2018. "Why the Repugnant Conclusion Is Inescapable." Working paper, Princeton CFI.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Budolfson, Mark, and Spears, Dean. 2019a. “Optimal Climate Policy Including Animal Welfare.” Working paper, Princeton CFI.

Budolfson, Mark, and Spears, Dean. 2019b. “Methods for Quantifying Animal Wellbeing and Estimating Optimal Tradeoffs against Human Wellbeing—And Lessons for Axiology, Including New Arguments for Separability.” Working paper, Princeton CFI.

Budolfson, Mark, and Spears, Dean. 2019c. “An Impossibility Result for Decision-Making under Normative Uncertainty.” Working paper, Princeton CFI.

Chan, Kai, Balvanera, Patricia, Benessaiah, Karina, Chapman, Mollie, Díaz, Sandra, Gómez-Baggethun, Erik, … Nancy Turner 2016. “Why Protect Nature? Rethinking Values and the Environment.” *Proceedings of the National Academy of Sciences* 113, no. 6: 1462–1465.

Connelly, Matthew. 2008. *Fatal Misconception*. Cambridge, MA: Harvard University Press.

Cowen, Tyler. 2006. “Market Failure for the Treatment of Animals.” *Society* 43, no. 2: 39–44.

Dasgupta, Partha. 2014. “Measuring the Wealth of Nations.” *Annual Review of Resource Economics* 6:17–31.

Dawkins, Marion. 2012. *Why Animals Matter: Animal Consciousness, Animal Welfare, and Human Well-being*. New York: Oxford University Press.

Deaton, Angus. 2013. *The Great Escape*. Princeton, NJ: Princeton University Press.

Drummond, Michael, Sculpher, Mark J., Claxton, Karl, Stoddart, Greg L., and Torrance, George W. 2005. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd ed. New York: Oxford University Press.

Ellerman, Denny Bailey, Elizabeth M., Montero, Juan-Pablo, Joskow, Paul, and Schmalensee, Richard L. 2000. *Markets for Clean Air*. Cambridge: Cambridge University Press.

Fischer, Robert. 2016. “What If Klein & Barron Are Right about Insect Sentience?” *Animal Sentience* 9, no. 8.

Fischer, Robert, Gamborg, Christian, Hampton, Jordan, Owen, Palmer, Clare, and Sandøe, Peter. Forthcoming. *Wildlife Ethics*. Malden, MA: Wiley.

Fleurbaey, Marc. 2015. “On Sustainability and Social Welfare.” *Journal of Environmental Economics and Management* 71:34–53.

(p. 613) Fleurbaey, Marc Ferranna, Maddalena, Budolfson, Mark, Dennig, Francis, Mintz-Woo, … Zuber, Stéphane. 2019. “The Social Cost of Carbon: Valuing Inequality, Risk, and Population for Climate Policy. *Monist* 102:84–109.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Fraser, David. 2008. *Understanding Animal Welfare*. Malden, MA: Wiley.

Garnett, Tara et al. 2013. “Sustainable Intensification in Agriculture: Premises and Policies.” *Science* 341, no. 6141: 33–34.

GiveWell.org. 2019. “Overview.” **https://www.givewell.org/about/givewell-overview**.

Gruen, Lori. 2011. *Ethics and Animals*. Cambridge: Cambridge University Press.

Hausman, Daniel, McPherson, Michael, and Satz, Debra. 2018. *Economic Analysis, Moral Philosophy and Public Policy*. 3rd ed. Cambridge: Cambridge University Press.

Herculano-Houzel, Susan. 2017. “Numbers of Neurons as Biological Correlates of Cognitive Capability.” *Current Opinion in Behavioral Sciences* 16:1–7.

Hiller, Avram. 2014. “System Consequentialism.” In *Consequentialism and Environmental Ethics*, edited by Hiller et al. New York: Routledge.

Hiller, Avram, Ilea, Ramona, and Kahn, Leonard, eds. 2014. *Consequentialism and Environmental Ethics*. New York: Routledge.

Hsiung, Wayne, and Sunstein, Cass. 2007. “Climate Change and Animals” 155 *University of Pennsylvania Law Review*. 1695.

IPBES. 2019. “Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.” **https://www.ipbes.net/**.

IPCC. 2014. “United Nations Intergovernmental Panel on Climate Change.” **https://www.ipcc.ch/**.

Jamieson, Dale. 2008. *Ethics and the Environment*. Cambridge: Cambridge University Press.

Jarvis, Lovell, and Donoso, Pablo. 2018. “A Selective Review of the Economic Analysis of Animal Health Management.” *Journal of Agricultural Economics* 69, no. 1: 201–225.

Kagan, Shelly. 2011. “Do I Make a Difference?” *Philosophy & Public Affairs* 39:105–141.

Kagan, Shelly. 2019. *How to Count Animals*. New York: Oxford University Press.

Kahneman, Daniel, and Deaton, Angus. 2010. “High Income Improves Evaluation of Life but Not Emotional Well-being.” *Proceedings of the National Academy of Sciences* 107, no. 38: 16489–16493.

Klein, Colin, and Barron, Andrew. 2016. “Insects Have the Capacity for Subjective Experience.” *Animal Sentience* 9, no. 1.

Kolstad, Charles. 2010. *Environmental Economics*. 2nd ed. New York: Oxford University Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Lam, David. 2011. “How the World Survived the Population Bomb: Lessons from 50 Years of Extraordinary Demographic History.” *Demography* 48, no. 4: 1231–1262.

MacAskill, Will, Bykvist, Krister, and Ord, Toby. Forthcoming. *Moral Uncertainty*. New York: Oxford University Press.

McMahan, Jeff. 2002. *The Ethics of Killing*. New York: Oxford University Press.

McShane, Katie. 2018. “Why Animal Welfare Is Not Biodiversity, Ecosystem Services, or Human Welfare.” *Les Ateliers de l’Éthique/The Ethics Forum* 13, no. 1: 43–64.

Mendl, Michael, and Paul, Elizabeth. 2016. “Bee Happy: Bumblebees Show Decision-Making That Reflects Emotion-Like States. *Science* 353:1499–1500.

Mikati, Ihab Benson, Adam F., Luben, Thomas J, Sacks, Jason D, and Richmond-Bryant, Jennifer. 2018. “Disparities in Distribution of Particulate Matter Emission Sources by Race and Poverty Status.” *American Journal of Public Health* 108, no. 4: 480–485.

Nefsky, Julia. 2011. “Consequentialism and the Problem of Collective Harm: A Reply to Kagan.” *Philosophy & Public Affairs* 39:364–395.

Nefsky, Julia. 2019. “Collective Harm and Individual Inefficacy.” *Philosophy Compass* 14, no. 4: e12587.

(p. 614) Ng, Yew-Kwang. 1995. “Towards Welfare Biology: Evolutionary Economics of Animal Consciousness and Suffering.” *Biology and Philosophy* 10:255–285.

Nilsson, Mans, Griggs, Dave, and Visbeck, Martin. 2016. “Policy: Map the Interactions between Sustainable Development Goals.” *Nature* 534:320–322.

Nolt, John. 2011. “How Harmful Are the Average American’s Greenhouse Gas Emissions?” *Ethics, Policy and Environment* 14, no. 1: 3–10.

Norcross, Alastair. 2004. “Puppies, Pigs, and People.” *Philosophical Perspectives* 18:229–245.

Nordhaus, William. 2007. “A Review of the Stern Review on the Economics of Climate Change.” *Journal of Economic Literature* 45, no. 3: 686–702.

Norwood, F. Bailey, and Lusk, Jayson. 2011. *Compassion by the Pound*. Oxford: Oxford University Press.

Olkowicz, Seweryn Kocourek, Martin, Lučan, Radek K., Porteš, Michal, Tecumseh Fitch, W., … Pavel Němec. 2016. “Birds Have Primate-like Numbers of Neurons in the Forebrain.” *Proceedings of the National Academy of Sciences* 113, no. 26: 7255–7260.

O’Neill, John, Holland, Allan, and Light, Andrew. 2008. *Environmental Values*. New York: Routledge.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

OPP. 2019. “Open Philanthropy Project, Farm Animal Focus.” **https://www.openphilanthropy.org/focus/us-policy/farm-animal-welfare**.

Ostrom, Elinor. 1990. *Governing the Commons*. Cambridge: Cambridge University Press.

Otte, M. Joachim, and Chilonda, Pius. 2000. “Animal Health Economics, in Livestock Information.” Rome, Italy: Sector Analysis and Policy Branch, Animal Production and Health Division (AGA), FAO.

Palmer, Clare, McShane, Katie, and Sandler, Ronald. 2014. “Environmental Ethics.” *Annual Review of Environment and Resources* 39:419–442.

Parfit, Derek. 1984. *Reasons and Persons*. New York: Oxford University Press.

Portmore, Douglas. 2011. *Commonsense Consequentialism*. New York: Oxford University Press.

Sandler, Ronald. 2018. *Environmental Ethics*. New York: Oxford University Press.

Sarkar, Sahotra. 2012. *Environmental Philosophy*. Malden, MA: Wiley-Blackwell.

Schmidtz, David, and Shahar, Dan. 2018. *Environmental Ethics*. 3rd ed. New York: Oxford University Press.

Scovronick, Noah Budolfson, Mark B., Dennig, Francis, Fleurbaey, Marc, Siebert, Asher, … Fabian Wagner. 2017. “Impact of Population Growth and Population Ethics on Climate Change Mitigation Policy.” *Proceedings of the National Academy of Sciences* 114, no. 46: 12338–12343.

Sebo, Jeff. Forthcoming. “Animals and Climate Change.” In *Philosophy and Climate Change*, edited by M. Budolfson, T. McPherson, and D. Plunkett. New York: Oxford University Press.

Shrader-Frechette, Kristin. 2002. *Environmental Justice*. New York: Oxford University Press.

Shriver, Adam. 2014. “The Asymmetrical Contributions of Pleasure and Pain to Subjective Well-Being.” *Review of Philosophy and Psychology* 5: 135–153.

Singer, Peter. 1972. “Famine, Affluence, and Morality.” *Philosophy & Public Affairs* 1, no. 3: 229–243.

Singer, Peter. 1975. *Animal Liberation*. New York: HarperCollins.

Singer, Peter. 1980. “Utilitarianism and Vegetarianism.” *Philosophy & Public Affairs* 9:325–337.

Singer, Peter. 2016. “Are Insects Conscious?” *Project Syndicate*.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Solow, Robert. 1991. *Sustainability: An Economist's Perspective*. The 18th J. Seward Johnson Lecture. Woods Hole: Woods Hole Institute.

Stiglitz, Joseph, Sen, Amartya, and Fitoussi, Jean-Paul. 2009. "Report by the Commission on the Measurement of Economic Performance and Social Progress." Government of France.

Sunstein, Cass. 2014. *Valuing Life*. Chicago: University of Chicago Press.

(p. 615) Sunstein, Cass. 2018. *The Cost-Benefit Revolution*. Cambridge, MA: MIT Press.

Tännsjö, Torbjörn. 2016. "It's Getting Better All the Time." In *Food, Ethics, and Society*, edited by A. Barnhill, M. Budolfson, and T. Doggett. New York: Oxford University Press.

Tye, Michael. 2017. *Tense Bees and Shell-Shocked Crabs*. Oxford University Press.

Wong, Kevin. 2016. "Counting Animals: On Effective Altruism and the Prospect of Interspecies Commensurability." PhD thesis, Princeton University.

Notes:

([1]) It is possible to endorse a different decision theory. On normative grounds, this has been advocated for different reasons by Buchak (2013) and Portmore (2011); thus, it is important to be explicit that 3 is a further assumption independent of 1 and 2. Because they are extensively discussed elsewhere, we will not focus on these alternatives here. We also set aside the important issue of decision-making under normative uncertainty—for discussion, see MacAskill et al. (forthcoming) and Budolfson and Spears (2019c).

([2]) In contrast, when the anthropocentric value of animals are well-reflected in market prices—such as e.g., the price of pollination services—market prices are the preferred method of valuation, at least to the extent that the good is a private good traded in a well-functioning marketplace.

([3]) Compare the concept of *enlightened anthropocentrism* in Sandler (2018).

([4]) Similarly, anthropocentrism assigns no fundamental value to the health of ecosystems, which is a different criticism—see the references in previous sentence, and in addition Chan et al. (2016) and Dasgupta (2014) and the references therein.

([5]) Compare also the term $\overline{f}_s$ to McMahan (2002)'s concept of *fortune*, a connection Wong (2016) notes.

Mark Budolfson

Mark Budolfson is Assistant Professor in Population-Level Bioethics, Philosophy, and Environmental Health Sciences at Rutgers University. He works on issues in philosophy, politics, and economics. Current research includes global ethics and internation-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

al institutions, population-level bioethics, sustainable development and climate change economics, and reasons for action in collective action situations.

Dean Spears

Dean Spears is an economic demographer and development economist. His research areas include the health, growth, and survival of children in developing countries and population dimensions of social well-being. Dean is Assistant Professor of Economics at the University of Texas at Austin, is a visiting economist at the Economics and Planning Unit of the Indian Statistical Institute in Delhi, is a founding Executive Director of r.i.c.e. (a nonprofit that works for children's health in India), and is an affiliate of IZA, of the Institute for Futures Studies, and of the Climate Futures Initiative at Princeton University. With Diane Coffey, he is a coauthor of the book Where India Goes: Abandoned Toilets, Stunted Development, and the Costs of Caste (2017); he is the author of Air: Pollution, Climate Change, and India's Choice between Policy and Pretence (2019). His research is supported by an NIH Population Scientist career grant.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The Science of Effective Altruism

Victor Kumar

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.27

Abstract and Keywords

This chapter evaluates effective altruism and its link to science. Contrary to much philosophical discussion, effective altruism is not tied essentially to utilitarianism and therefore does not suffer from the criticisms directed against utilitarianism. Prominent criticisms of effective altruism itself are unconvincing, since they identify remediable problems within the surrounding social movement and not problems essential to the theory itself. As a philosophical theory, effective altruism is worthy of allegiance because it strengthens a laudable connection between moral decision making and scientific evidence. Some philosophers and scientists believe that science supports utilitarianism, but their arguments are unpersuasive. Effective altruism is most plausible when it is divorced from utilitarianism. Despite that, effectice altruism can still encourage large and positive changes to our moral practices.

Keywords: effective altruism, utilitarianism, science, structural injustice, special obligations, debunking

PETER Singer might be the most influential moral philosopher since Plato. Among the general public, his work is immensely popular (at least when compared with other philosophical work in the analytic tradition). Almost single-handedly, Singer (1975) has sparked a revolution in attitudes toward nonhuman animals. Many people are vegetarians or vegans (or activists) principally because of him. Singer (2009; 2015a) has also made a powerful case for charitable giving toward people suffering from poverty and disease in the developing world. He's helped people grasp "the most good they can do."

Singer's career has followed in the tradition of other utilitarian social reformers like Jeremy Bentham and John Stuart Mill. These utilitarians responded to some of the most pressing social problems of their day. Bentham (1789) and Mill (1863) opposed racist and sexist discrimination. Their ideas are unremarkable nowadays, perhaps, but went decidedly against the grain of moral thought in nineteenth-century Europe. Singer, by contrast, has focused on indifference toward nonhuman animals and the global poor. At the same time, Singer has arguably also been undersensitive to the needs and interests of people with cognitive disabilities (see, e.g., Barnes 2016). His commitment to utilitarianism may have

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

engendered certain moral errors (cf. Taylor 2017). Nonetheless, like Bentham and Mill, Singer has led the way on addressing important forms of social injustice, transcending society's existing limits.

Singer has not just influenced people who read his work or hear him speak. He has also sparked a growing and cohesive philosophical position and social movement called "effective altruism." The movement is led by younger philosophers-cum-philanthropists like William MacAskill, Toby Ord, and others. These effective altruists follow in Singer's footsteps but aspire to leave an even bigger moral footprint.

Effective altruism is grounded in the idea that we should do the most good we can by using our resources in the most effective ways possible. If altruism is to be truly "effective," it must be guided by scientific evidence about how to make our money and our efforts go the furthest. Instead of picking what seems intuitively like the worthiest cause, (p. 532) one that imparts to the giver a "warm glow," effective altruists argue that it is better to donate to international charities that will save the greatest number of lives per dollar. One controversy surrounding effective altruism, as we'll see, stems from the possibility that global development charities offer only drops in a bucket. The bigger problems, allegedly, are systemic. Charitable giving doesn't address these systemic problems, according to some critics, and may well have the unintended effect of sustaining unjust systems.

Effective altruism puts utilitarian ethical theory into practice. Its aim is to develop and enact *a science of doing good*, thereby integrating utilitarianism with science. However, Singer (2005) also claims that utilitarianism itself is scientifically justified. This is different from using science to apply utilitarian principles, as effective altruists intend. According to Singer, science supports utilitarian principles. This is a bold idea, to say the least, in part because of the well-known gap that divides "is" from "ought." Among philosophers, the idea that science supports utilitarianism is often met with extreme skepticism, if not scorn. Academics in other disciplines are often more receptive, however. Utilitarianism seems to enjoy broad appeal among scientists, or at least among those scientists who give thought to moral philosophy. This is a curious fact, indeed so curious that it begs for explanation.

My main aim in this chapter is to evaluate the scientific merits of effective altruism and utilitarianism. I'll construct a defense of effective altruism that identifies promise in its commitment to revise particular moral commitments on the basis of scientific evidence. I'll also argue, however, that effective altruism is more plausible when stripped of its supposedly utilitarian foundations. Science, alas, offers no support to utilitarianism itself. But the evidential role for science in effective altruism endows this philosophical position and social movement with significant moral and intellectual worth.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

1. Utilitarianism and Effective Altruism

Utilitarianism is a beguiling ethical theory, as many ethics instructors can attest. One common argument for utilitarianism begins by asking what, if anything, has inherent value. Of all the many things people pursue, which are worthy in and of themselves, apart from their instrumental benefits? Why seek a large income? Why strive to be successful in your chosen profession? Why cultivate connections to friends and family? A tempting thought is that each of these pursuits is valuable insofar as it increases pleasure and alleviates suffering. Pleasure and the absence of suffering can be lumped together as "happiness." Thus, imagine that the connection between your pursuits and happiness is severed. If more money, greater professional success, or deeper relationships don't engender happiness, then it seems as though they are not worth pursuing.

Utilitarians suggest that simply by reflecting on your own experiences and activities, you can discover that happiness alone advances your interests. Other things advance (p. 533) your interests only indirectly—only by having an impact on pleasure or suffering. Once you see that happiness is so intimately connected to your own interests, it should then become clear that happiness matters just the same when others experience it. Pleasure cannot become more valuable when you enjoy it than when I do, at least not from "the point of view of the universe" (Sidgwick 1907). So, to act morally, you must consider how your action affects not just your own interests but also the interests of others. What matters is the impact on everyone's happiness, not the fact that other people are distinct individuals, nor that they live in a different place or time, nor even that they are members of a separate biological species.

Utilitarianism thus has two main tenets. One is a principle of inherent value: only happiness directly advances a person's interests. The other main tenet is a principle of equal consideration: everyone's interests matter equally, so long as they are capable of experiencing pleasure or suffering. Put these two principles together and you get utilitarianism—or "classical" utilitarianism. In every choice, to wit, the morally correct action is the one that has the highest expected value in terms of overall net happiness, while weighing the interests of every sentient being affected.

This is certainly a very brief presentation of utilitarianism and just one argument for it. We haven't considered other principles that are less salient but also essential to the theory. For example, classical utilitarianism is also committed to the principle of "maximizing" expected value rather than "satisficing" (cf. Lichtenberg, Chapter 29, this volume). In light of this principle, morally correct actions should produce the most expected happiness, rather than simply more or "enough." We also haven't yet considered objections to utilitarianism. Some of the most pressing objections draw out implications of the theory for rights and partiality that are hard to swallow.

However, it isn't the purpose of this chapter to fully flesh out utilitarianism or offer a thorough evaluation of its merits. My central focus will be on effective altruism (hereafter "EA"). Over the course of this chapter I will show how utilitarianism leads to EA and ex-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

plain what this view entails, in theory and in practice (this section). Next, I will defend EA from objections (sections 2 and 3). I will then argue that EA need not be grounded in utilitarianism (section 3). And finally, I will lay out why utilitarianism does not enjoy the empirical support that some proponents advertise (section 4). Utilitarianism entails EA, but EA does not entail utilitarianism—that is, as I'll argue, given a version of EA that most merits allegiance.

Singer's work is the impetus for EA, but younger scholars inspired by Singer's work have recently taken the reins. One of the most prominent effective altruists is William MacAskill. In *Doing Good Better*, MacAskill (2015, 11) unpacks EA as follows:

> Effective altruism is about asking, "How can I make the biggest difference that I can?" and using evidence and careful reasoning to try to find an answer. It takes a scientific approach to doing good. Just as science consists of the honest and impartial attempt to work out what's true, and a commitment to believe the truth whatever that turns out to be, effective altruism consists of the honest and impartial attempt to work out what's best for the world, and a commitment to do what's best, whatever that turns out to be.

(p. 534) Remember that EA is both a philosophical position and a social movement. As a philosophical position, EA says that we should take "a scientific approach to doing good." That is, if we have the time and if available evidence is not already decisive, we should rely on scientific evidence to determine how we can do the most good. (Whether EA says that this exhausts our moral obligations will be discussed later in section 3.) As a social movement, EA is a collection of individuals and organizations committed to following this principle. In this essay, "EA," unless qualified, denotes the philosophical position. However, I will sometimes distinguish the position from the movement, theory from practice. For example, I will argue that some criticisms of the practice do not undermine the theory; I will also suggest how the practice might be reformed so as to better live up to the theory.

Utilitarianism, it seems, straightforwardly entails EA. The entailment holds given the background assumption that, as seems correct, science offers the best evidence about the impact of one's actions on the interests of others. MacAskill (2015, 12) thus argues that an effective altruist should seek scientific answers to a number of particular questions. For example, who will be impacted by your actions and by how much? What are alternative courses of action and what would their effects be? What are the chances that your action will succeed in its aims? Scientific tools can address these questions, giving answers that are needed to best apply the utilitarian principles of inherent value and equal consideration.

Like many other effective altruists, MacAskill's main focus is on the science of doing good in the developing world. This focus isn't arbitrary. People in developed nations are members of the "global 1 percent" (2015, 15–25). Thus, in light of the diminishing marginal utility of wealth, global charity is much more effective than charity within developed nations. In terms of what some economists and medical ethicists call "quality-adjusted life years" (QALYs) "[t]he same amount of money can do [roughly] one hundred times as

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

much to benefit the very poorest people in the world as it can to benefit typical citizens of the [developing world]" (22). Unless one happens to be a multi-millionaire, a charitable donation can only be relatively small—a mere drop in the bucket. However, "[i]t's the size of the drop that matters, not the size of the bucket, and, if we choose, we can create a [relatively] enormous drop" (25).

Having laid out the theoretical underpinnings of EA, MacAskill goes on to canvass relevant scientific evidence. For example, he outlines which international aid charities have an established record of cost-effective work (2015, 121–127). These include GiveDirectly, which offers cash transfers to poor households in Kenya and Uganda, no strings attached; Development Media International, which produces radio programs that educate families about basic health measures in several African countries; and the Against Malaria Foundation, which provides bed nets that protect people from mosquitoes and malaria in sub-Saharan Africa. All of these charities have empirically demonstrable efficacy and can scale up their work on the basis of further donations.

MacAskill also argues that scientific evidence leads to a number of counter-intuitive conclusions. Just as a science of the material world overturns many widely held factual positions, a science of doing good overturns many widely held moral positions. For (p. 535) example, MacAskill criticizes a number of ways in which people engage in "ethical" consumption of consumer products. Some people in wealthy countries favor purchasing sugar, coffee, and other foods only through "fair trade" companies that provide employees with a decent wage. But since it is only companies in relatively well-off countries that can afford to adopt fair trade practices, MacAskill argues, the result is that workers in relatively disadvantaged countries lose employment. Better, then, to buy cheaper products that aren't fair trade and donate the difference in price to effective charities (2015, 132–135). MacAskill also criticizes certain "green" lifestyle choices, like turning off lights, refusing to use plastic bags, and buying locally sourced goods. Empirical evidence suggests that each choice is relatively ineffectual. Much more effective is carbon offsetting. Instead of consuming products that entail fewer greenhouse gas emissions, it is better to buy cheaper products and subsidize projects that reduce carbon emissions elsewhere (135–140).

MacAskill goes on to argue that EA provides a useful framework for choosing a career (MacAskill 2015, chap. 9). If you want to make a difference, the most effective career is not necessarily in the nonprofit sector, even if the company you choose to work for happens to be very effective. If the person who would have been hired instead of you is nearly as skilled, then you aren't making much of a difference (155). MacAskill argues that one possible option is "earning to give" (163–164). If you have the right skills, it might be better to enter the financial sector instead of the world of nonprofits, earn a much higher salary, and give away a large proportion of your income to charities that will do more good than you could have accomplished through your own nonprofit work.

So far in this chapter, I've outlined one argument for utilitarianism, defined EA, and shown how utilitarianism entails EA as a philosophical position. I've also described EA in

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

practice by recounting some of the recommendations that seem to issue from a science of doing good. In the rest of the chapter I'll turn to the business of critically evaluating EA. I'll begin by evaluating EA on its own terms before returning to its connection with utilitarianism. We'll have to wait until the final main section of the chapter before we consider scientific arguments for utilitarianism and evaluate their merits.

2. The Structural Objection

Casual assessments of EA are sometimes anchored in particular recommendations given by effective altruists, like those recommendations rehearsed at the end of the previous section. Thus, for instance, some people who have a positive opinion about EA may defend it with an example. Surely it's a good idea to donate to the Against Malaria Foundation instead of wasting money on Toys for Tots? However, an apparently sensible recommendation like this is not necessarily what comes to mind for critics. People who have a negative opinion about EA sometimes appeal to a different example. Isn't it a bad idea to become a hedge-fund trader, instead of working directly with disadvantaged communities, simply because you would then have more surplus income available to donate to charities? Might this not be a betrayal of your values?

(p. 536) It is tempting to scrutinize particular recommendations given by effective altruists. Perhaps, for example, there is evidence that people who enter the financial sector are likely to forget about the noble ambitions that first led them there. A science of doing good must, of course, be open to this possibility and therefore willing to abandon this recommendation (cf. MacAskill 2015, 165–167). But this just shows that to assess EA as a philosophical theory we must go beyond examples of the practice. We can't let applications of the theory stand in for the theory as a whole.

One recurrent criticism of the practice of EA lies in its focus on advancing the interests of people in the developing world through financial aid. Many people are intensely skeptical about the value of international aid organizations. Some aid programs are ineffectual or worse. In addition, some programs are captured by the interests of corrupt institutions, whether in the countries where aid originates or in the countries where aid is targeted. However, MacAskill (2015, 47–53) claims that it's a mistake to focus on the worst cases. Overall, he argues, international aid has been an enormous success. It suffices that the best organizations have been extremely effective. MacAskill thus argues that the eradication of smallpox alone is worth all the money spent on aid programs (46). Arguments against donating to bad aid programs do not count against donating to good aid programs.

A deeper criticism of EA lies underneath the surface here—which I will call the "structural objection." The objection begins by arguing that suffering in the developing world stems from massive global inequalities in power. These inequalities rest on local and international institutions and the hierarchical social structure they engender. Effective altruists like MacAskill want to work with these institutions, instead of reforming them. According to the structural objection, however, the root of suffering in the developing world

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

is systemic and structural inequality that gives wealthy nations outsized power and control in international affairs. This type of inequality cannot be opposed through financial aid, but only through political action, not individually but collectively. Drops, even enormous ones, will have limited value if the bucket is lopsided by design.

A number of critics have given voice to this objection. Amia Srinivasan (2015) expresses it lucidly in her gripping and wide-ranging review of MacAskill's book:

> Effective altruism, so far at least, has been a conservative movement, calling us back to where we already are: the world as it is, our institutions as they are. MacAskill does not address the deep sources of global misery—international trade and finance, debt, nationalism, imperialism, racial and gender-based subordination, war, environmental degradation, corruption, exploitation of labor—or the forces that ensure its reproduction. Effective altruism doesn't try to understand how power works, except to better align itself with it.

Srinivasan is more sympathetic to EA than some critics. She shares with effective altruists a similar perspective about the world's moral problems and accords them a similar priority. But she argues that the structural or systemic bases of the problems cannot be addressed through charity (see also, e.g., Herzog 2016).

(p. 537) Jeff McMahan (2016) does not find the structural objection quite so compelling. His response to the objection begins by pointing out that a moral agent does not have direct influence over social and political institutions. What she can directly control are her own actions and efforts. She may attempt to reform global economic institutions or she may take direct charitable action. It is best to do both (McMahan 2016):

> Yet there has to be a certain division of moral labor, with some people taking direct action to address the plight of the most impoverished people, while others devote their efforts to bringing about institutional changes through political action. To suppose that the only acceptable option is to work to reform global economic institutions and that it is self-indulgent to make incremental contributions to the amelioration of poverty through individual action is rather like condemning a doctor who treats the victims of a war for failing to devote his efforts instead to eliminating the root causes of war.

Let's linger on McMahan's last point.

In "Famine, Affluence, and Morality," Singer (1972) famously asked readers to imagine the following case. You are walking by a shallow pond and see a toddler drowning. You can rush in, but doing so will ruin your very expensive new clothes. Are you obligated to save the child? Of course! Having taken the bait, however, you are now on the hook. In fact, you have the opportunity to save a child's life in the developing world for a financial cost that is similar (very roughly) to the price of your expensive clothes. The child is remote instead of nearby, starving instead of drowning, but the differences between the

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

cases do not seem to be morally relevant. Every day that you choose material luxuries over charity you are effectively refusing to save a drowning child.

Provided that Singer's analogy is sound, proponents of the structural objection seem to be reasoning as follows: "You are walking by a shallow pond and see a child drowning. Are you obligated to save the child? No! Not if the child is drowning because of deeper structural causes, say, because the local government refuses to erect barriers around the pond. If that's the case, you should leave immediately and lobby the government to make structural changes, leaving the child to die in favor of more important problems." Effective altruists are sometimes held in suspicion for taking a heartless, mechanistic approach to ethics. However, it now seems to be their critics, with an eye on long-term structural change, who lack appropriate moral concern.

Still, perhaps it is true that more good can be done through political activism than through charity. Perhaps, in other words, you should sometimes let drowning children drown. Utilitarians, for one, are open to counter-intuitive moral implications such as this. Recall that MacAskill (2015, 11) says we should "do what's best, whatever that turns out to be." Suppose, then, for the sake of argument, that suffering in developing countries is the result of structural factors, that structural reform is necessary to alleviate suffering, and that collective political action rather than individual charity has greater priority. Even then, however, the practice might need to be reformed, but EA would yet remain viable. The structural objection, too, doesn't go deep enough to evaluate the theory itself. It speaks (p. 538) only to application of the theory. As Singer (2015b) himself says, "[e]ffective altruism cannot be refuted by evidence that some other strategy will be more effective than the one effective altruists are using, because effective altruists will then adopt that strategy."

EA is the view that we should take a scientific approach to doing good. We should not rely on unreflective judgments about which causes are worthy. Given the moral problems at issue, and given that the solutions are not obvious, we should seek empirical evidence about which solutions are most likely to address them. This principle is broad enough to apply not just to how we spend surplus income but also to how we apportion our political efforts. Thus, some organizations aligned with EA, like The Humane League, focus on structural reform to advance animal welfare (see Weathers 2016). For projects like this, empirical evidence would seem to be necessary. We should not simply devote ourselves to forms of political activism that leave us with a warm glow. We need, in short, not just a science of charity but also a science of political activism. Some philosophers, like Jeff Sebo, have recently taken up the project of studying animal welfare activism (Sebo and Singer 2018; Sebo 2019).

To see more clearly how EA can absorb the structural objection, consider the following dilemma. On the one hand, suppose that there is empirical evidence that political activism does the most good, for example, activism that seeks to reform global economic institutions. This might be measured in terms of QALYs or perhaps according to some other measure. Then, by its own lights, EA should accord activism greater priority than charity.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

On the other hand, suppose that empirical evidence does not show that activism does the most good. Then EA should continue to prioritize charity. But in that case, it seems, any moral agent should think so, too. That is, a responsible moral agent should respect scientific evidence about the effects of political activism instead of simply going with their gut. Doing otherwise seems epistemically irrational. As McMahan says, recall, activism versus charity is not an either/or question. But there are still questions of priority and resource allocation, for example, the proportion of effective altruists devoted to either activity. EA's science of doing good provides a useful framework for thinking about the social division of moral labor within the movement.

If EA can be reconciled with the structural objection in the way I've suggested, however, further amendments to the usual practice of EA are required. As a theory, effective altruism is committed to a science of doing good. In practice, though, effective altruists tend to be highly selective with respect to scientific evidence. They rely mainly on economics, especially development economics, as relevant to "how to do good better." Economics is well positioned to study the effects of charity where it is needed most. In practice, then, EA reflects a bias toward interventions that are relatively easy to measure through quantitative methods (Sebo and Singer 2018). In some ways this bias makes good sense. If we can't accurately measure the efficacy of an intervention, we seem to have less reason to support it. Nonetheless, in light of our discussion in this section, the sources of evidence that EA practitioners rely on must be expanded.

We need here to distinguish two types of questions: general and specific. The general question is about what *general causes* most deserve our money or efforts. For example, (p. 539) MacAskill argues that global poverty is one area in which people can do a lot of good. One reason, recall, is that charitable contributions go so much further in poor countries than in the developed world. But MacAskill also thinks other causes are similarly worthy, including factory farming, climate change, immigration, and criminal justice reform (2015, 185–193). Each of these general causes, he argues, involves social problems that are large in scale, neglected, and potentially tractable. More recently, MacAskill has pursued "longtermism," trying to think seriously about how to improve the lives of people thousands or millions of years into the future (supposing that humans make it that far).

General questions about worthy causes lead to more specific questions. Holding fixed the general cause, specific questions remain about *which programs or actions* are the best way of aiding the cause. Thus, for example, MacAskill argues that deworming programs in the developing world have a higher impact on educational outcomes than programs that donate textbooks to schoolchildren (2015, 104–108). Much of MacAskill's book is devoted to using economics to answer various specific questions. When it comes to the area of political activism, however, it's not clear that economics is well positioned to answer specific questions about *how to pursue political activism* (at least given the current state of the discipline). Other scientific fields can potentially offer more insight. For example, work in sociology attempts to study what types of political movements are most effective.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Consider the work of the sociologist Charles Tilly (Tilly and Tarrow 2015; Tilly 2006; see Anderson 2014 for further discussion). Tilly claims that social movements are effective when they can publicly demonstrate the possession of four features: (1) apparent worthiness; (2) unity among the members; (3) high number of adherents; and (4) commitment in the face of personal sacrifice. This type of scientific research is absent from existing EA practice, but it would seem to be crucial if people are to make decisions about how to effectively participate in political activism—not whether to participate in activism at all or what priority it has (general question), but which political activities are most worthwhile (specific question). Without relying on research like this, it seems, effective altruists cannot carry out a sufficiently broad and robust science of doing good.

3. Effective Altruism without Utilitarianism

I've argued that EA has the ability to meet what is perhaps the most serious criticism leveled against it. However, even if you agree that the structural objection fails in the end, you may still wonder what positive reason there is in favor of EA in the first place. If you think there is no reason to accept EA, you may regard objections and responses as internal disputes that have no wider relevance. Of course, utilitarianism entails EA, but that is no reason to accept EA if there are strong reasons to reject utilitarianism.

(p. 540) In this section, I'll argue that EA need not rest on utilitarianism (see also McMahan 2016; MacAskill 2017). To make this case, I'll continue to rely on an interpretation of EA as a science of doing good. That is, EA is the view that we should rely on scientific evidence to determine how we can do the most good. But I'll argue explicitly now that, given the arguments offered for EA, it does not rule out the existence of other moral obligations, including obligations for which science is not particularly relevant. So the version of EA that I'll defend does not line up with everything that utilitarians like Singer and MacAskill say about it. However, I'll argue that this makes EA more plausible, not less. The result is a version of EA "worth wanting" and yet one that still imposes rather strong demands on moral agents. Some effective altruists, indeed, might welcome a utilitarianism-free version of EA if it were then, as a consequence, to gain wider appeal and spur more philanthropy and thus be more fruitful on utilitarian grounds.

Let's begin with another persistent objection to EA (aside from the structural objection), one that stems from Bernard Williams's (1973) famous critique of utilitarianism. I won't try to fully unpack Williams's ideas or attempt to be precisely faithful to them, but I hope to say enough to explain why one might still be drawn to EA, or something like it, even if one is sufficiently persuaded by Williams's critique to reject full-blooded utilitarianism.

Williams argues that utilitarianism is unattractive because it is too impersonal and ignores the "separateness of persons" (see Brink, Chapter 20, this volume). It treats individuals as mere instruments of utility maximization and ignores their partial perspectives and commitments (see Jeske, Chapter 12, this volume). According to utilitarianism, doing the most good is the only thing that matters, no matter the particular goals and projects

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

that a person has and that seem to make her life meaningful. For these reasons, utilitarianism is "alienating."

As McMahan (2016) observes, many critics of EA channel Williams, directly or indirectly attacking EA's allegedly utilitarian foundations. For example, Nakul Krishna (2016) insists that there is value in the supposed "hokeyness involved in the business of finding ourselves and our deepest impulses." Ethics cannot and should not eliminate personal cares from its purview. What one cares about means something:

> [EA presents a] picture of moral reflection as arbitration between the claims of different people, one of whom just happens to be me. In this picture, it seems like the fact that *I'm me* has been declared, right at the outset, irrelevant. To direct my charitable donations to training guide dogs for the blind (an obscenely inefficient way of doing good, the effective altruists say) would be to treat (mistakenly) the fact that I happen to care about this cause as if it meant something.

Suppose you are persuaded by the thought that ethics must not treat a person's cares and projects as morally irrelevant. What does this mean for EA? I'll argue presently that it casts doubt on some applications of EA, but that EA itself remains intact.

Williams-style reservations about utilitarianism clash with some of the recommendations that effective altruists make, in particular, those that ignore partiality. MacAskill (2015, 41–42; emphasis added) considers partiality and rejects its moral significance:

(p. 541)

> If I were to give to [a foundation with which I have a personal connection] rather than to the charities that I thought were most effective, I would be privileging the needs of some people over others merely because I happened to know them. That would be *unfair* to those I could have helped more. … For example, if an uncle dies of cancer, you might naturally want to raise money for cancer research. Responding to bereavement by trying to make a difference is certainly admirable. But it seems *arbitrary* to raise money for one specific cause of death rather than any other. … By all means, we should harness the sadness we feel at the loss of a loved one in order to make the world a better place. But we should focus that motivation on preventing death and improving lives, rather than preventing death and improving lives in one very specific way. Any other decision would be *unfair* to those whom we could have helped more.

Now, it's possible that many people will have enhanced motivation to donate to charities when they have personal connections to them, possible then that yielding to personal connections will do more good overall than aiming to be most efficient. Were there empirical evidence supporting these ideas, MacAskill and other effective altruists might grant that partiality has moral significance, though only indirect. (On the other hand, motivational dispositions are not to be taken as fixed and can themselves be the target of influence.) However, scrutinizing this example again does not go deep enough to evaluate EA as a

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

philosophical position. Our concern is with the idea that it is “arbitrary” or “unfair” to privilege personal connections over impersonal interests. This can’t be convincing to anyone persuaded by Williams (cf. Gabriel 2018).

Utilitarianism rejects the very idea of “special obligations,” that is, obligations that one has not to sentient or sapient beings in general but in virtue of personal relationships. We might have special obligations in virtue of our personal cares and commitments, or in virtue of duties of reparations for past wrongs. Williams gives us reasons to hold on to these obligations. Nonetheless, even if special obligations are not eliminated, any plausible moral theory will have to include so-called general obligations that one has to others in general. These include *obligations of beneficence*. Such obligations could but need not be cashed out in terms of happiness. Moreover, it might yet be true that we should fulfill obligations of beneficence through impartial charity or activism, that donating money or time to special causes does not suffice to fulfill these obligations, and that most of us should donate much more than we currently do. EA thus can still have bite, Williams notwithstanding.

It’s striking that in MacAskill’s book he nowhere mentions utilitarianism. Nor does he argue for EA on utilitarian grounds, that is, along the lines rehearsed in section 1 of this essay. Like Singer himself, MacAskill makes his case for EA by appealing to moral intuitions that are widely held and seemingly quite plausible. For example, MacAskill (2015, 30–32) describes the process of medical triage and the decisions doctors and nurses must make to prioritize those patients who can most benefit from treatment. Medical professionals must make these difficult decisions, leaving some patients to suffer or die, in order to save others. Decisions about charitable contributions are similar to triage, MacAskill argues. They involve making “hard trade-offs” (32). Or consider MacAskill’s (p. 542) (23) argument for channeling financial aid to developing countries, given that money goes roughly 100 times further there than in the developed world:

> It’s not often that you have two options, one of which is one hundred times better than the other. Imagine a happy hour where you could buy yourself a beer for five dollars or buy someone else a beer for five cents. If that were the case, we’d probably be pretty generous—next round’s on me! But that’s effectively the situation we’re in all the time. It’s like a 99-percent-off sale. … It might be the most amazing deal you’ll see in your life.

The strongest argument for EA is thus not that it follows from utilitarianism. And it doesn’t lead to a view on which our personal projects are morally insignificant. Rather, EA claims that a science of doing good is the best way to fulfill our obligations of impartial beneficence and that we are required to fulfill these obligations in the best way possible. The strongest argument for this view is that it follows from ordinary, plausible intuitions to which we already seem to be committed, like the intuition Singer evokes in his famous example. Intuitively, one is obligated to save a child drowning in a nearby pond and there is no morally relevant difference in the case of starving children. Intuitively, as well, we should save more children rather than fewer. But it remains an open question

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

whether, in any given case, obligations of beneficence trump special obligations. Without an all-encompassing theory like utilitarianism, such questions can never be closed and will always depend on the details.

I've now argued that EA is best construed as a theory about how to fulfill certain general obligations—in particular, obligations of beneficence—not a consequence of the utilitarian worldview that eschews special obligations. That is, EA says that if we have the time and if available evidence is not already decisive, we should rely on scientific evidence to determine how we can do the most good so as to fulfill obligations of impersonal beneficence. Note that I've simply assumed that there is something to Williams's criticism of utilitarianism, without considering possible responses. The ultimate merits of this criticism, I grant, are not obvious. But my point is that even if you relinquish utilitarianism, there are still powerful reasons to hang on to EA.

Nonetheless, from this perspective, some applications of EA are no longer supported. For example, it may be permissible to give something to charities with which you have a personal connection. In addition, it does not follow that you should choose a career that enables "earning to give"—not if pursuing a meaningful life also matters. However, EA still offers plenty of other recommendations that a morally responsible agent should act on. A science of doing good is ineliminable if we have a responsibility, as it seems we do, to effectively combat such ills as global poverty, animal suffering, and climate change.

4. Scientific Utilitarianism

Up till now in this chapter, I've been critically reviewing the literature on effective altruism. I've argued that EA is attractive in that it says moral decisions should be grounded (p. 543) in scientific evidence. This virtue of EA can be preserved even if those sympathetic to the view prioritize collective political action over individual financial charity. The virtue can also be preserved even if we abandon a totalizing utilitarianism and maintain a commitment to special obligations and meaning via personal cares and commitments. However, it has been argued by Singer himself, among others, that a scientifically grounded ethics leads, after all, to utilitarianism. In that case, we could avail ourselves of a simpler and more straightforward argument for EA. And the amendment to EA given in the previous section—that it applies only to some of our obligations—would turn out to be unnecessary.

Why would one think that science leads to utilitarianism? A fairly simple-minded reason is that scientists themselves tend to be utilitarians—that is, if they subscribe to any moral theory at all. I'll argue in a moment that, as one might expect, this is not a very good reason for thinking that science supports utilitarianism, not even the best one. But the mere fact that utilitarianism appeals to scientists is interesting enough to merit attention. I'll lay out a couple of explanatory hypotheses for this sociological phenomenon, but I won't spend time defending them since I want to turn to a better scientific argument for utilitarianism, and contend that this argument doesn't hold water either.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

One reason many scientists are drawn to utilitarianism might be simply that if offers a mathematical approach to ethical decision-making that is reminiscent of mathematical approaches in science. Utilitarianism asks a moral agent to precisely specify the outcomes of actions, along with their probabilities, in order to calculate the expected value of each possible action. Then, to figure out the morally correct choice, it seems, one simply has to do the math. Thus, a mathematical approach that is sensible in empirical domains may strike scientists as sensible in the ethical domain, too. However, while this might be what causes some scientists to be utilitarians, it isn't a very good reason—not without some argument for thinking that ethics is relevantly like science. Mathematical approaches are not sensible, perhaps not even intelligible, in domains like aesthetics or literature. Furthermore, as Tyler John (personal communication) points out in this context, nonutilitarian and even nonconsequentialist moral theories can be formalized, too, in ways that can appeal to mathematically oriented minds (see Hurley, Chapter 2, this volume for relevant discussion of "consequentializing"). As a result, this explanation for utilitarianism's popularity might debunk, rather than vindicate, scientists' commitment to the theory (see Kumar 2017).

I suspect there is an additional reason that scientists are drawn to utilitarianism. The idea of instrumental reasons (or instrumental value) is relatively clear, even to those who lack philosophical training. Instrumental reasons would seem to fit quite naturally within the materialist worldview favored by scientists. To a first approximation, instrumental reasons consist in relations of cause and effect. That something is instrumentally valuable is, or seems to be, just the fact that it helps to bring about a good outcome. Because utilitarianism embodies a minimal commitment to intrinsic or inherent value—it does not truck with rights or justice or any other noninstrumental sources of value or reasons aside from happiness—it appeals to people who are friendly toward instrumental reasons and wary of other normative categories. Herein, I believe, lies an argument that, were it to be unpacked more fully, is potentially quite powerful (which (p. 544) is not to say decisive). That is, scientific materialism might cast doubt on the idea of noninstrumental reasons. A thorough commitment to this argument would lead to moral nihilism rather than utilitarianism. However, even given other reasons to reject nihilism, the argument would support only consequentialism generally and not utilitarianism specifically (see Hubin 2001 for relevant discussion). Whatever it is that has noninstrumental value need not be happiness.

Though they remain only hypotheses, for all that I have said, we have on the table now two explanations for why scientists tend to believe in utilitarianism. However, since these explanations don't vindicate the beliefs (Kumar 2017), we don't yet have an argument for why science supports utilitarianism. We'll turn next to one such argument that is widely discussed in the literature. It is fueled not by economics or sociology but by cognitive science and evolutionary biology.

Utilitarianism is beguiling in its simplicity. It also has quite radical and demanding implications (see Sobel, Chapter 11, this volume). Utilitarianism seems to entail that individuals should sacrifice all of their personal interests for the sake of others. Members of the

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

global 1 percent should not simply funnel some small portion of their resources to the developing world. Given the way in which money goes so much further there, utilitarianism entails that those of us living relatively comfortable lives should donate all of our money until there is no person worse off than us whose interests can yet be advanced. Or, depending on the evidence and the math, we should devote all of our time to collective political action and give up everything else that presumptively makes our lives meaningful.

Many philosophers have thought that utilitarianism is untenable because of other implications that are not just radical but deeply counter-intuitive. Utilitarianism denies the existence of rights and justice that transcend happiness, not to mention the moral significance of personal cares and commitments. For these reasons, it seems to violate plain moral common sense, or even moral decency. However, Singer (2005) claims that although utilitarianism conflicts with widespread moral intuitions, these intuitions are not trustworthy given their evolutionary and psychological origins. Our moral intuitions are the product of the very particular ecological circumstances that gave rise to the human moral mind (cf. Kumar and Campbell, unpublished manuscript). As a consequence, Singer (2005, 348) argues, moral common sense is not to be trusted:

> There is little point in constructing a moral theory designed to match considered moral judgments that themselves stem from our evolved responses to the situations in which we and our ancestors lived during the period of our evolution as social mammals, primates, and finally, human beings. We should, with our current powers of reasoning and our rapidly changing circumstances, be able to do better than that.

Singer is arguing for utilitarianism indirectly, by offering an evolutionary and psychological debunking argument against the intuitions that seem to undermine it. This debunking argument has been developed in more detail by the philosophically trained psychologist Joshua Greene (2007; 2014a; 2014b), who is also responsible for some of the empirical research that fuels it. Let’s focus on his argument.

(p. 545) Greene argues that evolutionary forces gave rise to certain “moral heuristics.” These heuristics have a number of cognitive and motivational consequences. They lead us to be partial to our friends and family, since this was crucial in small-scale communities of hunter-gatherers. They also make us averse to harming others through direct “personal force,” even when more good can be produced that way, since this aversion was likewise essential to cooperation in the Pleistocene.

In the “environment of evolutionary adaptedness,” then, the heuristics that underlie moral intuitions had survival value. Greene suggests that they also likely produced the most good in that environment, relative to feasible alternatives. Nowadays, however, in large-scale, technologically advanced societies, these heuristics lead us astray, according to Greene. Our current environment makes partiality harmful and creates plenty of opportunities through which people can be harmed indirectly without “personal force,” including through inaction. Moral intuitions are thus somewhat like the psychological drives that led our ancestors to cash in on rare sources of fat and sugar when times were

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

lean, as they often were in the Pleistocene, but that lead to obesity in many modern environments.

The problem with Greene's arguments, in essence, is that he must rely on the claim that moral intuitions are driven largely by morally irrelevant factors. This claim is not substantiated. For example, Greene does not seem to possess an argument that is independent of utilitarianism, and thus that isn't question-begging, for thinking that partiality is morally irrelevant. Greene's critics should grant that whether or not harm is inflicted through personal force is morally irrelevant. This claim is plausible, but it doesn't go nearly far enough, as I've argued elsewhere (Kumar 2017, 125–126):

> [Moral] intuitions are sensitive to a range of [other] factors. … For example, intuitions track the degree of harm inflicted, whether it was caused intentionally or only accidentally, whether it was intended as a means to an end or merely as a foreseen side effect, whether it was a deserved response to aggression or unprovoked, and so on. … Greene must [claim] that these *other factors* influencing intuition also do not lend rational credibility to … them. The problem for Greene is that this … is not at all plausible.

Greene explicitly does not attempt to leap from "is" to "ought." He offers a debunking argument that rests not just on empirical claims about human psychology but also on normative claims about what is and is not morally relevant. In general, a debunking argument is successful insofar as its normative premises are more plausible than the normative claims it attempts to debunk (Kumar and Campbell 2012). By these standards, Greene's argument is unsuccessful, given the implausibility of the normative claim that all or most of the factors that drive moral intuition are morally irrelevant.

My criticisms of Greene and Singer, like their own arguments, rest on empirical research in cognitive science. More detailed articulation of these criticisms and closer readings of the empirical evidence can be found elsewhere (see Kumar and Campbell 2012; Kumar 2017; Kumar and May 2019). Here I want to end by suggesting a new argument that is in some sense "a priori" in that it doesn't rest on evidence from cognitive science.

(p. 546) Greene argues that the psychological mechanisms that underlie people's moral intuitions are faulty. The rationally innocent mechanisms are those that underlie their (conflicting) commitment to utilitarianism. For the sake of argument, suppose we grant that "the rational parts of ourselves" are drawn to utilitarianism. But for all that cognitive science says, that may be because utilitarianism is deceptively plausible, because it takes in rational minds with its alluring but shallow simplicity. Science can tell us what draws people to utilitarianism, even rationally, but it cannot tell us whether the theory is credible in the end. Only philosophy can do that.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

5. Summary

Science doesn’t support utilitarianism. But science does play a valuable role in EA. We need a science of doing good no matter what broader ethical theory we subscribe to, even if we subscribe to no broad ethical theory at all. There are two main objections to EA that I’ve argued do not win the day. If there is good evidence that we can best ameliorate global poverty and suffering through collective political activism that seeks structural reform, then EA should recommend that. If utilitarianism fails to eliminate special obligations that arise from our personal cares and commitments, then EA isn’t our only ethical guide but still provides a science of how to fulfill general obligations of beneficence. Singer, MacAskill, and other effective altruists offer persuasive arguments—grounded in intuitions about concrete cases and not in broad ethical theories—that people in developed countries must do more and do better. This isn’t all that effective altruists seek to establish, but it is more than good enough.

AcknowledgmentsI’m grateful to Samia Hesni, Tyler John, Judith Lichtenberg, Meghan Nesmith, Douglas Portmore, and Aja Watkins for very helpful comments on previous drafts. Work on this chapter was supported by the Peter Paul Professorship at Boston University.

References

Anderson, Elizabeth. 2014. “Social Movements, Experiments in Living, and Moral Progress: Case Studies from Britain’s Abolition of Slavery.” The Lindley Lecture, The University of Kansas.

Barnes, Elizabeth. 2016. *The Minority Body: A Theory of Disability*. Oxford: Oxford University Press.

Bentham, Jeremy. 1789. *An Introduction to the Principles of Morals and Legislation* Oxford: Oxford University Press.

Gabriel, Iason. 2018. “The Problem with Yuppie Ethics.” *Utilitas* 30, no. 1: 32–53.

Greene, Joshua. 2007. “The Secret Joke of Kant’s Soul.” In *Moral Psychology*, vol. 3, edited by W. Sinnott-Armstrong, 35–80. Cambridge, MA: MIT Press.

(p. 547) Greene, Joshua. 2014a. *Moral Tribes: Emotion, Reason, and the Gap between Us and Them*. London: Penguin Books.

Greene, Joshua. 2014b. “Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)science Matters for Ethics.” *Ethics* 124, no. 4: 695–726.

Herzog, Lisa. 2016. “Can “Effective Altruism’ Really Change the World?” *Open Democracy*, Feb. 22.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Hubin, Donald. 2001. “The Groundless Normativity of Instrumental Rationality.” *Journal of Philosophy* 98, no. 9: 445–468.

Krishna, Nakul. 2016. “Add Your Own Egg.” *The Point Magazine*, January 14.

Kumar, Victor. 2017. “Moral Vindications.” *Cognition* 167:124–134.

Kumar, Victor, and Campbell, Richmond. Unpublished manuscript. “A Better Ape: How Morality Drives Human Evolution.”

Kumar, Victor, and Campbell, Richmond. 2012. “On the Normative Significance of Experimental Moral Psychology.” *Philosophical Psychology* 25, no. 3: 311–330.

Kumar, Victor, and May, Joshua. 2019. “How to Debunk Moral Beliefs.” In *Methodology and Moral Philosophy*, edited by Jussi Suikkanen and Antti Kauppinen, 25–48. New York: Routledge.

MacAskill, William. 2015. *Doing Good Better: How Effective Altruism Can Help You Make a Difference*. New York: Gotham Press.

MacAskill, William. 2017. “Effective Altruism: Introduction.” *Essays in Philosophy* 18, no. 1: 1–5.

McMahan, Jeff. 2016. “Philosophical Critiques of Effective Altruism.” *The Philosophers’ Magazine* 73:92–99.

Mill, John Stuart. 1863. *Utilitarianism*. London: Parker, Son and Bourn.

Sebo, Jeff. 2019. “Effective Animal Advocacy.” In *The Routledge Handbook of Animal Ethics*. New York: Routledge.

Sebo, Jeff, and Singer, Peter. 2018. “Activism.” In *Critical Terms for Animal Studies*, 33–46. Chicago: University of Chicago Press.

Sidgwick, Henry. 1907. *The Methods of Ethics*. London: MacMillan and Co.

Singer, Peter. 1972. “Famine, Affluence, and Morality.” *Philosophy and Public Affairs* 1, no. 3: 229–243.

Singer, Peter. 1975. *Animal Liberation: The Definitive Classic of the Animal Movement*. New York: Harper.

Singer, Peter. 2005. “Ethics and Intuitions.” *Journal of Ethics* 9, no. 3–4: 331–352.

Singer, Peter. 2009. *The Life You Can Save: How to Do Your Part to End World Poverty*. New York: Random House.

Singer, Peter. 2015a. *The Most Good You Can Do: How Effective Altruism Is Changing Ideas about Living Ethically*. New Haven, CT: Yale University Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Singer, Peter. 2015b. “The Logic of Effective Altruism.” *Boston Review*, July 1.

Srinivasan, Amia. 2015. “Stop the Robot Apocalypse.” *London Review of Books*, September 24.

Taylor, Sunaura. 2017. *Beasts of Burden: Animal and Disability Liberation*. New York: The New Press.

Tilly, Charles. 2006. *Identities, Boundaries and Social Ties*. Boulder, CO: Routledge.

Tilly, Charles, and Tarrow, Sidney. 2015. *Contentious Politics*. New York: Oxford University Press.

Weathers, Scott. 2016. “Can “Effective Altruism’ Change the World? It Already Has.” *Open Democracy*, February 29.

Williams, Bernard. 1973. “A Critique of Utilitarianism.” In *Utilitarianism for and Against*, edited by J. C. C. Smart and Bernard Williams. Cambridge: Cambridge University Press.

Victor Kumar

Victor Kumar is Assistant Professor at Boston University. He works mainly at the intersection of ethics and cognitive science. His published work can be found in Ethics, Noûs, and Philosophers’ Imprint. In recent years he has written about moral learning, moral luck, and moral disgust. He is currently writing a book with Richmond Campbell about moral evolution and moral progress.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Effective Altruism: A Consequentialist Case Study

Judith Lichtenberg

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.28

Abstract and Keywords

In this essay I examine the contemporary movement known as effective altruism (EA). I argue that most understandings of EA imply some version of consequentialism. That in itself may sound like a rather modest conclusion (dictated by a certain vagueness in EA and the cornucopia of forms of consequentialism), but the arguments for it illuminate aspects of both EA and consequentialism. I also argue that the claim that one is obligated to maximize the good is not essential to consequentialism, that in fact this is a difficult claim to defend, and that therefore the standard "demandingness objection" misses the target. Nevertheless, what is essential to any consequentialist theory is the view that producing more good is always morally better than producing less. Deontological criticisms of this view are familiar. I focus instead on its clash with common-sense views about moral goodness and admirability.

Keywords: consequentialism, utilitarianism, effective altruism, demandingness, virtue

1.

EFFECTIVE altruism (EA) emerged over the last decade out of the ideas and writings of two young Oxford philosophers, Toby Ord and William MacAskill. In 2009, Ord founded Giving What We Can, an organization whose members pledge to donate 10 percent of their income to "whichever organisations can most effectively use it to improve the lives of others." As of June 2019, it had over 4,000 members who have pledged over $126,000,000 to charity.[1] MacAskill and Benjamin Todd started 80,000 Hours (the average number of hours a person works in her lifetime) in 2011 to help people figure out how to best use those hours "to solve the world's most pressing problems." They report that "More than 3000 people have told us that, due to engaging with us, they have significantly changed their career plans and now expect to have a larger social impact as a result."[2]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The movement has its roots in the works of Peter Singer, beginning with his classic 1972 essay “Famine, Affluence, and Morality,” a staple of introductory ethics courses ever since. Writing in the wake of a Bangladesh famine, Singer reset the course of contemporary moral philosophy—and certainly the teaching of it—with his deceptively simple premise that “if it is in our power to prevent something bad from happening, without thereby sacrificing anything of comparable moral importance, we ought, morally, to do it.”[3] With global poverty as his focus, Singer concludes without much ado that “we ought to give until we reach the level … at which, by giving more, I would cause as much suffering to myself or (p. 549) my dependents as I would relieve by my gift.” In so doing, Singer, an avowed utilitarian, married his philosophical theory to the so-called demandingness problem. Ever since, moral philosophers have been asking whether it is reasonable or morally right to expect people to make such significant sacrifices to alleviate the suffering of others.

Singer’s 2015 book *The Most Good You Can Do: How Effective Altruism Is Changing Ideas about Living Ethically* puts the point more clearly and explicitly. In contrast to the original, negative formulation in terms of “preventing something bad from happening,” it expresses the point positively—“Effective altruism is based on a very simple idea: we should do the most good we can.”[4] In his 2015 book *Doing Good Better: How Effective Altruism Can Help You Make a Difference*, MacAskill similarly asserts that EA is about making the biggest difference you can. “Of all the ways in which we could make the world a better place, which will do the most good?”[5]

Singer’s and MacAskill’s books, and other products of the EA movement, have undoubtedly generated some very good results. For one thing, they help rebut the scores of books and articles published over the last several decades arguing that international “aid” is at best ineffective and at worst counterproductive. These books have names like *The White Man’s Burden*, *Dead Aid*, *The Road to Hell*, *Lords of Poverty*, *Famine Crimes*, and *The Dark Sides of Virtue*.[6] Aid’s critics do not really deny that it’s possible for affluent westerners to improve the lives of the world’s poorest people—improving health outcomes is the most obvious example—but that fact often gets obscured by the titles and headlines. Effective altruists help set this record straight. They show how without enormous sacrifices people can greatly improve the odds that their donations will make substantial improvements in human well-being.

But EA, I shall argue, rests on problematic assumptions. Are effective altruists right that one should always do the most good one can? Even if the answer is no, is doing more good always morally better than doing less? Before tackling these questions I consider a prior one: does EA necessarily presuppose consequentialism?

(p. 550) 2.

I take consequentialism to be the view that the rightness or wrongness of acts depends solely on their consequences—their tendency to increase the good or decrease the bad, however good and bad are understood. Utilitarianism, the classic form of consequential-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ism, identifies the good with pleasure or happiness or well-being and the bad with its opposite. Utilitarianism has declined in popularity over the last few decades and has been replaced among consequentialists by more pluralistic accounts of the good.

Singer is an avowed consequentialist (a utilitarian, even), and many other prominent advocates probably are as well. Whether consequentialism is an essential feature of EA depends what one takes the latter's core ideas to be. If the point is simply to do more to address the world's greatest problems, such as global poverty, and to do it more effectively, the answer is no.[7] Over the last several thousand years, many religious, ethical, and political traditions have advocated remedying poverty and inequality. The Bible urges us to "Sell what you have and give the money to the poor."[8] Augustine asserts that "The superfluity of the rich is necessary to the poor. If you hold onto superfluous items, then you are keeping what belongs to someone else."[9] Liberals, socialists, Marxists, communists, and others have, on moral grounds, long called for a more equitable distribution of wealth that allows the least well off to live decently (if not equally). Even economists, invoking the principle of diminishing marginal utility, find good reasons to defend redistribution from richer to poorer (ceteris paribus). So the idea that those who are able to ought to do more, even much more, to alleviate poverty, and that we should not waste our efforts, are hardly unique insights of EA and do not entail consequentialism.

How, then, does EA differ from these age-old injunctions, if it does? The suspicion is that it presents us with an extremely demanding command to maximize the good (or minimize the bad), as its godfather, Peter Singer, has urged. The "demandingness problem" has beset contemporary ethics, and the maximizing version seems to be the default interpretation of consequentialism today.

In a brief essay sympathetic to EA, Jeff McMahan denies that the view presupposes consequentialism. Referring to Singer's argument in "Famine, Affluence, and Morality," McMahan says it "appealed in the first instance to a single widely held moral intuition and argued that consistency required those who accepted the intuition to give most of their wealth to the relief of extreme poverty."[10] Likewise for Peter Unger's argument for a similar conclusion in *Living High and Letting Die*, which, McMahan says, explicitly disavowed commitment to any particular moral theory. "His aim was to demonstrate that a (p. 551) view of the sort that now informs the work of effective altruists is implicit in values and convictions we already have."[11]

But do Singer's and Unger's conclusions follow from "a single widely held moral intuition" or from "values and convictions we already have"? Equally plausible is that once we understand what these philosophers think these values and intuitions imply, what appears to be a widely held moral intuition (such as "if it is in our power to prevent something bad from happening, without thereby sacrificing anything of comparable moral importance, we ought, morally, to do it") is one we are ready to discard. We see that we misunderstood this "intuition" and do not in fact accept it. That, at least, is what I take away from my decades of teaching Singer's article to thousands of undergraduates. To assert that our assent to Singer's principle on first and casual reading amounts to a widely held intu-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ition or deep-seated value is simply wrong.[12] On a pragmatic understanding of meaning, we do not accept this principle if we reject what seem to be its obvious implications.

So do effective altruists insist that we do the most good we can, as Singer's title implies? Or should we merely "do good better," as MacAskill's title suggests? If the latter, is that anything more than common sense? Amia Srinivasin puts the dilemma this way:

> Either effective altruism, like utilitarianism, demands that we do the most good possible, or it asks merely that we try to make things better. The first thought is genuinely radical, requiring us to overhaul our daily lives in ways unimaginable to most. … The second thought—that we try to make things better—is shared by every plausible moral system and every decent person."[13]

I doubt there is a single answer that applies to all who consider themselves effective altruists. For one thing, EA is at least as much a movement as a theory, and as such it seeks to cast a wide net, embracing people with different underlying philosophies. (Consequentialists, in fact, can recommend such a move, on the grounds that it produces better consequences by not setting the bar too high or making the membership criteria too strict.) Remember MacAskill's question: "Of all the ways in which we could make the world a better place, which will do the most good?"[14] This formulation suggests that (p. 552) making the world a better place is not mandatory, but that insofar as you adopt that goal you should make it as good as you can. But MacAskill also says that "Effective altruism consists of the honest and impartial attempt to work out what's best for the world, and a commitment to do what's best, whatever that turns out to be."[15] That sounds more like Singer's demanding view.

Julia Wise, president of Giving What We Can and community liaison for the Centre for Effective Altruism, defends a more moderate position. She asserts that although cost-effectiveness is a useful tool that people should apply to solving global problems, it is not applicable to all parts of a person's life:

> not everything that you do is in the "effectiveness" bucket. I don't even know what that would look like. … If I donate to my friend's fundraiser for her sick uncle, I'm pursuing a goal. But it's the goal of "support my friend and our friendship," not my goal of "make the world as good as possible." … That money is coming from my "personal satisfaction" budget, along with getting coffee with my friend. … I have another pot of money set aside for donating as effectively as I can. When I'm deciding what to do with that money, I turn on that bright light of cost-effectiveness and try [to] make as much progress as I can on the world's problems.[16]

Is it acceptable, then, to have different pots of money—some to satisfy our personal desires, some to support relationships we care about, some for solving global problems like poverty and disease—with amounts allocated just as you please? Should Wise's donation to the fundraiser for her friend's sick uncle be understood simply in terms of "personal satisfaction," not so different from getting coffee with a friend or buying a new dress? What about the pot of money "set aside for donating as effectively as I can"? Well, if the

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

pot is for donating as effectively as you can, it's tautological that with it you should donate as effectively as you can.

Presumably effective altruists do not mean to state a tautology, so let's ask what nontautological recommendation might follow from their statements. One might be to be careful, when donating to causes like global poverty, to make sure the organization you support is not wasteful, corrupt, or inefficient, and that it is making good use of your donation. This is pretty uncontroversial, hardly enough to propel a movement meant to inspire people, as EA seems to be. Is it wrong to give to an organization that is not the most effective? Or, if not wrong, is it less morally good? How should a person decide (p. 553) how much to allocate to different causes, some of which are more effective than others? Effective altruists aim to persuade people to give much more than they currently do to organizations that effectively address global poverty (and other large problems). And this inevitably leads us to ask how demanding their view is.

3.

It's an interesting question why demandingness was not an issue for consequentialists until around the last third of the twentieth century—spurred no doubt in part at least by Singer's article. John Stuart Mill expressed a much more moderate view in *Utilitarianism*:

> The multiplication of happiness is, according to the utilitarian ethics, the object of virtue: the occasions on which any person (except one in a thousand) has it in his power to do this on an extended scale—in other words, to be a public benefactor—are but exceptional; and on these occasions alone is he called on to consider public utility; in every other case, private utility, the interest or happiness of some few persons, is all he has to attend to.[17]

Henry Sidgwick took a similar line, asserting that "each person is for the most part, from limitation either of power or knowledge, not in a position to do much good to more than a very small number of persons."[18] The world Mill and Sidgwick inhabited was very different from ours, it's true; the efficacy of fairly ordinary people today may well be greater than it was in their day. But perhaps also they were self-deceived, or at least negligent in their thinking.

Mill softens the demandingness problem in his very definition of utilitarianism: "actions are right *in proportion* as they tend to promote happiness, wrong as they tend to produce the reverse of happiness."[19] Today this is often called scalar consequentialism: the more good you do the better; but it doesn't follow that you are morally *obligated* to do the most good.[20] Utilitarians can draw lines (on utilitarian grounds, of course) at places that evaluative terms like "wrong," "impermissible," "blameworthy," and the like (including their positive counterparts) should mark.[21]

(p. 554) Sidgwick, by contrast, argues that "a Utilitarian must hold that it is always wrong for a man knowingly to do anything other than what he believes to be most conducive to

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Universal Happiness."[22] G. E. Moore thought similarly, writing, in *Principia Ethica*, that "the assertion 'I am morally bound to perform this action' is identical with the assertion 'This action will produce the greatest amount of good in the Universe.'" Moore insists that the point is "demonstrably certain."[23]

The maximizing view became standard. In 1979, Brian Barry described it as "the time-bomb that has been ticking away ever since" and that "has at last blown up utilitarianism."[24] But Barry was wrong; maximizing consequentialism is alive and well. For example, Shelly Kagan, a leading consequentialist, takes consequentialism to mean that "Agents are morally required to perform the act that will lead to the best results overall."[25]

Neither Singer nor MacAskill uses the term "duty" or "obligation" in these works, nor do they say you are wrong if you do not do what's best; rather, they say you *ought* to do the best thing you can. "Ought" is an ambiguous word. Perhaps, then, they are not maximizing consequentialists but instead accept Mill's scalar view. If so, doing less than the best is not necessarily wrong. Nevertheless, Singer asks us to do the most good we can, and other effective altruists suggest similar things. In at least one place Singer implies that one who does not aim to bring about the very best consequences is morally deficient and open to criticism.[26] So there is a crucial ambiguity running through the pronouncements of effective altruists.

But unless it is so defined, there is nothing in consequentialism that entails the maximizing view. "Consequentialism … is simply the view that normative properties depend only on consequences."[27] "A moral theory is a form of consequentialism if and only if it assesses acts and/or character traits, practices, and institutions solely in terms of the (p. 555) goodness of the consequences."[28] Consequentialism entails that the more good you do the better, but it is not part of its meaning that you are morally required to do what is best. And I find no convincing arguments by consequentialists for the conclusion that maximizing the good is morally required. (As I will suggest later below, it's hard to know what would count as a persuasive argument for it.) Where we should draw the lines between acts that are prohibited, permissible, praiseworthy, and the like will depend, according to consequentialism, on what promotes the best consequences.

Nonmaximizing consequentialism is defanged of a property that makes consequentialism highly controversial. Still, essential to consequentialism is the idea that it is always morally *preferable* to bring about better consequences than worse ones.

This view has often been challenged on deontological grounds, by those who believe it is sometimes *wrong* to bring about more good rather than less. But there are other reasons to doubt it too.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

4.

Consider Jane, a chemistry professor in her forties with a successful career and a good salary in a respected university. She gives generously to causes devoted to alleviating global poverty. She does not impoverish herself in the process, but she donates much more than the average American household and acts in line with recommendations of the EA movement, carefully researching nongovernmental organizations (or relying on others, like GiveWell, to do so) and giving her money to organizations with a proven track record of success in alleviating the worst problems facing the global poor.

But over the years Jane has developed interests in other major social problems too. In particular, she has become deeply disturbed by the glaring social and economic inequalities in the United States and the legacy of racism and discrimination underpinning them. So disturbed is she, in fact, that she eventually decides to quit her academic job to work with at-risk teenagers in her city, mostly members of minorities—working to keep them in school, off drugs, out of trouble, and on track to go to college or learn a trade. She also works part-time for a nonprofit organization with similar goals. Together, these activities amount to a full-time job and then some. But she makes much less than what she made as an academic, significantly diminishing the resources she has to give to global (p. 556) poverty relief. Over the twenty or more years she would likely have continued as an academic, those foregone contributions, according to MacAskill's calculations, could have saved some significant number of lives. Jane's work saves no lives (that we know of). And let's suppose that despite the good it achieves it does not reduce human suffering as much as those donations would have (however we choose to measure these things).

Many would consider Jane admirable, but by the standards of EA she is open to criticism. Consider what the organization 80,000 Hours advises to those who wish to relieve human suffering. First, figure out which problems are "large in scale, solvable and neglected." As of this writing, the organization lists "positively shaping the development of artificial intelligence," "biorisk reduction," nuclear security, and climate change.[29] Health in poor countries is also a priority, although it seems to have fallen a few notches in the last several years, perhaps because it is less neglected as a field than some of the others on the list.[30] Effective altruists also note that global health and poverty alleviation will always take priority over their domestic analogues, largely because dollars going to the least well-off produce more utility than dollars going to those higher on the socioeconomic scale. The poorest 5 percent of Americans are "at the 68th percentile of the world income distribution."[31]

80,000 Hours' next recommendation is to engage in research, or government or policy work, in one of those needed areas; or to work at an effective nonprofit; or to "apply an unusual strength to a needed niche."[32] Another strategy is "earning to give"—taking a high-paying job, then donating a good chunk of one's salary to an urgent problem. In his book MacAskill describes Greg Lewis, an idealistic doctor who decided not to practice medicine in a poor country but instead to become a medical oncologist in the United Kingdom so that he could donate half his $200,000 earnings to global poverty relief.[33]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Singer begins his book with a discussion of Matt Wage, a brilliant former student of his at Princeton who declined to continue in philosophy despite an offer from Oxford and instead went to work for an arbitrage trading firm on Wall Street in order to be able to donate six-figure sums to highly effective charities.[34]

Many criticisms can and have been brought against EA, attempting to show its methods are not "better." That it is elitist, concentrating almost exclusively on how highly educated people can make a difference (the movement's home is in Oxford, and it shows); that it is individualistic and apolitical, whereas the problems it addresses are (p. 557) political and must be solved collectively by governments and other groups; that poverty is best addressed by economic development; that the techniques of EA are paternalistic and undemocratic, treating those it helps as passive and irrelevant.[35] To such criticisms Singer responds: "Effective altruism cannot be refuted by evidence that some other strategy will be more effective than the one effective altruists are using, because effective altruists will then adopt that strategy."[36] If political action is a better means to improving welfare than individual giving, that's what we should promote. If private investment by Western conglomerates works best, do that. But there are, it seems, no intrinsic advantages of these approaches.

This response can make it difficult for criticism of EA to stick. Is it an adequate answer? Let me explain why I believe it is not.

5.

One reason has to do with how much we can pack into the conception of the good without trivializing consequentialism. If we reject the simpler utilitarian conception—the good is pleasure, happiness, or well-being; the bad is pain, unhappiness, or ill-being—and add a bunch of other goods such as autonomy, dignity, integrity, self-determination, fidelity, democracy, and human rights, then we will need some way of weighing the various values against each other in cases where they conflict, as they inevitably will some of the time, to decide which takes priority in what circumstances. For consequentialism to be a distinctive moral theory there must ultimately be a single scale along which to measure these different values. Otherwise it resembles an intuitionism that judges (nonscientifically or nonquantitatively) which value takes precedence when.[37]

Effective altruists and consequentialists seem highly optimistic about the prospects for measuring, quantifying, and comparing these values—including the less hedonistic ones—to produce the right answer about what to do. Of course, effective altruists are rational, scientific-minded people who embrace fallibilism; even in the few years since the movement started they have changed their minds quite a bit about what strategies to pursue. Nonetheless, they seem to engage in predictions with great confidence. For example, among GiveWell's most recommended charities (there are only eight as of this writing) are Evidence Action's Deworm the World Initiative and the Schistosomiasis Control Initiative, both of which seek to eradicate intestinal worms, which afflict more (p. 558) than two billion people in poor countries.[38] But in 2015, Cochrane, an organization that evaluates

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

health and medical research data, cast doubt on the efficacy of mass deworming programs; its findings were confirmed in a report published in the *Lancet* in 2017.[39] Angus Deaton, a Nobel Prize–winning economist who is critical of EA, attributes the movement's overconfidence in this and other cases to overreliance on randomized experiments, which "consider only the immediate effects of the interventions, not the contexts in which they are set. Nor, most importantly, can they say anything about the wide-ranging unintended consequences."[40]

James Lenman describes the problem decision makers face as "massive and inscrutable causal ramification."[41] Thus, he argues, the kinds of predictions on which consequentialists rely and judge others' behavior are highly questionable. Consequentialists may respond that this is not a problem faced by them alone: everyone, whatever their theories, must make decisions under uncertainty. But consequentialists face the problem in a more severe form, since for them consequences are the *only* things that matter and *all* the consequences matter. Nothing else counts. GiveWell, for example, whether it explicitly embraces consequentialism or not, is vulnerable to these charges partly because it generally recommends that people put all their charitable eggs in just a few baskets. Less confidence might lead one to spread one's donations more widely.

6.

But let's leave this difficulty aside. The main problem I want to address arises from the fact that even a nonmaximizing, scalar consequentialist must be committed to the view that bringing about more good, whatever that good is, is always morally better than bringing about less. That is inherent in consequentialism. So, for example, what Jane does is worse than if she had gone to work for one of the organizations 80,000 Hours recommends or if she had "earned to give" (assuming that either or both of these would have produced better overall results). Her chosen course may be better than doing nothing, but it is far less morally good than other options she could have taken.

(p. 559) The question is whether this view, which I believe is at odds with our commonsense beliefs, can be made convincing by means other than building it into the definition of consequentialism. From a consequentialist point of view, what Greg and Matt do just *is* morally better than what Jane does. Here's another story that MacAskill tells. In 2009, while germinating the concept of EA, he visited a hospital in Ethiopia that treats obstetric fistulas, a condition resulting from childbirth in young and malnourished women that causes "permanent incontinence of urine and/or feces" and pariah status. According to the Fistula Foundation, which funds the hospital, "A majority of women who develop fistulas are abandoned by their husbands and ostracized by their communities because of their foul smell." At the hospital MacAskill met some of the women who suffered from this condition. But several years later he concluded that although the organization was repairing fistulas at low cost and saving these young women from terrible fates, others working on different issues were making a bigger impact and that they should get his donations instead. (The methods used in determining impact are described in the book in detail.)

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

MacAskill explains that by donating to the Fistula Foundation instead of a different organization he thought he "would be privileging the needs of some people over others merely because I happened to know them," and that it "was arbitrary that I'd seen this problem close up rather than any of the other problems in the world."[42]

Effective altruists often applaud MacAskill's approach, admiring the cool rationality that considers personal attachments arbitrary. Singer tells us that "many of the most prominent effective altruists have backgrounds in or are particularly strong in areas that require abstract reasoning."[43]

Effective altruists are right that people are often led astray, in a variety of ways, by their emotions and personal attachments, and that these can lead to pernicious biases. But critics may nevertheless find MacAskill's approach chilling. As Larissa MacFarquhar puts it, effective altruists fail to understand "that, to many people, to suppress emotional connection to make way for a more rational altruism is to crush their moral roots."[44]

There are several different worries implicit here. Philosophers have long criticized consequentialism on the grounds that it ignores the moral value of partiality. Even on the scalar version, it is always *permissible*, never wrong, to bring about more good rather than less, and that can mean neglecting those near and dear in favor of strangers. But, as Diane Jeske argues elsewhere in this volume, "many of us cannot help but think that choosing to benefit our friend [rather than a stranger] is not only morally permissible but, in fact, (p. 560) morally required."[45] Jeske, like many others, defends "partialism," according to which it is "(not merely psychologically understandable but) morally correct to favour one's own … [i.e.] those to whom the agent has some special relationship or personal tie."[46]

Two features of the kinds of examples offered by partialists are important. First, they involve cases in which one confronts the choice to benefit (or prevent harm to) an intimate rather than a stranger. Second, the partialist casts the matter in deontic terms: we have a *duty* to favor those with whom we have a relationship; it would be wrong not to.[47]

Jane's case exhibits neither of these features. Jane has no prior relationship with the children she works with; she makes a decision ex ante to work with them rather than to "earn to give." And, partly for that reason, it would be odd to say she is required to work with them rather than do something more utility-producing. In these cases, deontic language does not describe our attitudes. It would be more natural to say that Jane, or what she does, is *morally admirable*. We think highly of the traits associated with her behavior. We esteem certain human characters and characteristics. And we may believe that Jane's behavior is *no less admirable than earning to give*.

Can effective altruists and consequentialists make sense of these beliefs? They may resist what seems to them like a naïve conflation of "does more good" with "is more admirable" or "is a good person." But consequentialists must interpret the judgment of human character, like everything else, in consequentialist terms.[48] The strategies they may use to incorporate the common-sense view that Jane is no less admirable than (what I shall call)

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

the philanthropist resemble those employed against deontologically minded critics. Just as consequentialists can argue that in the long run the world will be a better place if people take care of their own and are partial to those near and dear to them, they may say that the world will go better if some people choose local, hands-on efforts such as working with at-risk teenagers or illiterate adults, addressing prison reform or homelessness in the United States, reporting on gang violence, or working for the rights of transgender people. The character traits possessed by such people are admirable because they tend to produce good consequences.

Yet today the claim that from a consequentialist point of view local efforts might produce just as much good as global efforts seems implausible. (And, indeed, effective altruists explicitly deny it.) If we compare the numbers of people who suffer globally from poverty and disease to, say, the numbers of Americans who do, and the extent to which (p. 561) these problems could be remedied, the former will swamp the latter by a good margin.[49] Exactly so!, the effective altruists may say. This shows, they think, that those who earn to give *are* more admirable than those who work hands-on.

To avoid this conclusion, consequentialists might argue that the traits that lead ordinary people to be generous or self-sacrificing or to help people in their own communities represent traits that, over the long haul of human history, have produced highly beneficial consequences. We should not be short-sighted; the benefits of encouraging the local, hands-on approach, and of doing what you can with what you have, are greater than the comparison of the contemporary global and local poor would suggest.

7.

How long a time frame must a consequentialist consider, then, in recommending courses of action? This draws us near to the other approach consequentialists have taken to soften the clash between their views and the common-sense commitment to partiality and localism: to adopt some form of indirect consequentialism, such as rule consequentialism or motive consequentialism. Rule consequentialism, the variation that has been most extensively developed, says we should adopt those general rules adherence to which would maximize the best consequences. What is the life expectancy of a rule? Must we change the rules from those that may have worked for most of human history? In its pure form, rule consequentialism is a very different beast from act consequentialism that can lead to wholly different recommendations for action. It has a Kantian flavor, asking “What if everybody did that?” Despite its appeal, rule consequentialism is often thought to be, as Philippa Foot puts it, an “unstable compromise” whose very consequentialism can be called into question. One threat is that it “collapses” into act consequentialism.[50] In any case, I find no evidence that contemporary effective altruists accept indirect consequentialism.[51]

(p. 562) Another strategy consequentialists might adopt is to include traits like admirability and goodness as intrinsic goods to be included among all the other things we think are good. It might not be any more difficult, or any more necessary, for them to explain on

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

what grounds we include such traits than it would be for anyone else, whatever their moral theory. But this will not solve the problem, because these goods will almost certainly be swamped by others (full stomachs, long and healthy lives). Here again effective altruists and other consequentialists may embrace the consequentialist implication—what we might call the QED response—but others will remain unsatisfied.

One unfortunate upshot of the effective altruist's preference for big philanthropy and highly skilled careers is to downgrade the lives and works of those without the skills to attain such careers and unable to earn large sums of money to donate. They might give generously, given their means, but their means severely limit their impact. Can consequentialists avoid the embarrassing conclusion that such people are less admirable than rich people who have a much greater impact?

We can assume they will not want to say that. Like the rest of us, they are probably drawn to the intuitive idea that one's "goodness" (admirability, virtue, decency, moral compass) is largely a function of what one does with what one's got, how much one does for others, and so on. But it is not clear how to incorporate any of these convictions on consequentialist grounds.

8.

It's not news that pure consequentialism conflicts with various of our common-sense beliefs. Its clash with deontological intuitions has dominated discussions of moral theory for decades. Here I have focused on a less well-trodden conflict: with our judgments about what constitutes moral virtue and admirability.

Effective altruists and other consequentialists are right to insist that the mere fact that our "intuitions" tell us that people like Jane are as good as the generous capitalist is no argument at all. If common sense were the last word, then slavery would have been justified to those who thought its legitimacy was intuitively clear.

But common-sense beliefs about morality are not always suspect. Common sense is especially untrustworthy when what it endorses happens to coincide with our interests. When it justified slavery to slaveholders, there were deep reasons to distrust it. Moreover, there were many reasons for thinking slavery was immoral. But no self-interested motive is apparent when common sense tells us that we should not knowingly convict the innocent or that civil rights workers are admirable. In such cases we have much less reason to distrust our intuitions.

Of course, our self-interest *is* at stake when we reject a theory that demands or even just recommends that we do more than would be convenient for us. Yet many moral and political theories, outlooks, and religions express ideals, aspirations, or even moral requirements that far surpass what most people do. What makes consequentialism (p. 563) unacceptable is not this, but rather its commitment to calculating and quantifying the Good,

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

which guarantees that many actions, traits, and habits that we value will be swamped by the numbers.

Consequentialists who admit the intrinsic value of traits like Jane's can recognize this outcome as an unfortunate but inevitable by-product of their theory. But some think it is enough to discredit any view that takes how the numbers fall out as decisive.

This disagreement cannot be rationally decided.

9.

Is EA necessarily consequentialist? Clearly, there are different versions of EA. No doubt not every effective altruist is a consequentialist—for one thing, some may never have heard of consequentialism or may not have a formally coherent moral outlook; some may simply think that those who can ought to do more for those in need. But of those who have thought in these terms I expect (and have argued here) that most are consequentialists, not necessarily in the maximizing sense but in the more modest sense of believing that producing more good is always morally better than producing less. And that in itself is controversial.

But EA is both a *movement* and a *theory* or *view*. Insofar as EA is a movement, its members will want to get other people to join. They may well decide that the best way to do that is by emphasizing aspects of the movement that are appealing, satisfying, and not too onerous—that do not much disrupt people's existing aims, values, and habits. Stern moral imperatives that challenge people's moral decency may be counterproductive. Effective altruists may well be committed to a theory, a set of moral truths—perhaps even that one ought to do as much as one possibly can to increase well-being overall—but may recognize that expressing the theory widely and publicly will not necessarily help to bring about its goals.[52] And effective altruists want, after all, to be effective.[53]

Notes:

([1]) Giving What We Can, https://www.givingwhatwecan.org/about-us/. In 2016, Giving What We Can was incorporated into its parent organization, the Centre for Effective Altruism in Oxford.

([2]) 80,000 Hours, at https://80000hours.org/. This organization is also sponsored by the Centre for Effective Altruism.

([3]) Peter Singer, "Famine, Affluence, and Morality," *Philosophy & Public Affairs* 1 (1972), 231.

([4]) Peter Singer, *The Most Good You Can Do: How Effective Altruism Is Changing Ideas about Living Ethically* (New Haven, CT: Yale University Press, 2015), vii.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([5]) William MacAskill, *Doing Good Better: How Effective Altruism Can Help You Make a Difference* (New York: Gotham Books, 2015), 32. In that sense the title of his book is misleading. A person might well do good *better* than she had but still not make the biggest difference she could. I discuss this ambiguity below.

([6]) See William Easterly, *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good* (New York: Penguin Press, 2006); Dambisa Moyo, *Dead Aid: Why Aid Is Not Working and How There Is a Better Way for Africa* (New York: Farrar Straus & Giroux, 2009); Michael Maren, *The Road to Hell: The Ravaging Effects of Foreign Aid and International Charity* (New York: Free Press, 1997); Graham Hancock, *Lords of Poverty: The Power, Prestige, and Corruption of the International Aid Business* (New York: Atlantic Monthly Press, 1989); Alex de Waal, *Famine Crimes: Politics and the Disaster Relief Industry in Africa* (London: Africa Rights and the International African Institute, 1997); David Kennedy, *The Dark Sides of Virtue: Reassessing International Humanitarianism* (Princeton, NJ: Princeton University Press, 2005). For further discussion of these critiques, see Judith Lichtenberg, *Distant Strangers: Ethics, Psychology, and Global Poverty* (Cambridge: Cambridge University Press, 2014), chap. 8.

([7]) I shall focus here on the alleviation of global poverty, one of EA's main concerns. It's not the only one, however, and not everything I say here will apply to all of the movement's targets.

([8]) Luke 12:33.

([9]) Augustine, *Exposition of the Psalms 121–150*, vol. III/20 (Hyde Park, NY: New City Press, 2004).

([10]) "Philosophical Critiques of Effective Altruism," http://jeffersonmcmahan.com/wp-content/uploads/2012/11/Philosophical-Critiques-of-Effective-Altruism-refs-in-text.pdf, 1.

([11]) Ibid., 1; see also Unger, *Living High and Letting Die: Our Illusion of Innocence* (Oxford: Oxford University Press, 1996).

([12]) Interestingly, in an earlier version of his essay McMahan writes: "my experience as a moral philosopher has been to find that common moral beliefs are often confused and inconsistent. … Whenever I consider a moral issue with care, I inevitably find the common sense view rather shallow. It is always possible to go deeper." This passage might be thought to support my claim: the superficially common-sensical view expressed in Singer's premise turns out on closer analysis to be confused and inconsistent. Of course, McMahan's statement was meant to support the opposite conclusion: "I therefore think it is a mistake to suppose that the moral views of effective altruists can be rejected on the ground that they are more demanding than people now and in the past have thought that morality could be." (Earlier version available from JL.)

([13]) Amia Srinivasin, "Stop the Robot Apocalypse," *London Review of Books* 37, no. 18, September 24, 2015, at http://www.lrb.co.uk/v37/n18/amia-srinivasan/stop-the-robot-apocalypse.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([14]) MacAskill, *Doing Good Better*, 32.

([15]) MacAskill, *Doing Good Better*, 11.

([16]) Julia Wise, "You Have More Than One Goal, and That's Fine," *Giving Gladly*, http://www.givinggladly.com/, February 19, 2019. Peter Singer writes about Wise in *The Most Good You Can Do*, 23–31, as does Larissa MacFarquhar in *Strangers Drowning: Grappling with Impossible Idealism, Drastic Choices and the Overpowering Urge to Help* (New York: Penguin Press, 2015), 71–87. MacFarquhar shows a different aspect of Wise: since childhood she had "believed that because each person was equally valuable she was not entitled to care more for her own well-being than for anyone else's" (73). Perhaps she has made peace with these more radical and guilt-inducing views over the years. It's also possible that, as an official representative of EA organizations, she realizes the need to encourage people to adopt its ways without inducing undue guilt.

([17]) J. S. Mill, *Utilitarianism*, 2nd ed., edited by George Sher (Indianapolis: Hackett, 2001), 19.

([18]) Henry Sidgwick, *The Methods of Ethics* (New York: Dover, 1966) (republication of the 7th ed., 1907), 434. Sidgwick here seems not recognize the possibility of culpable ignorance.

([19]) Mill, *Utilitarianism*, 7. Emphasis added.

([20]) Terms like "scalar consequentialism" are often technically defined. I am using the term in a broad sense, to include theories that draw no lines but just denote degrees (of goodness or badness), as well as those that do draw lines of the sort I describe.

([21]) Although I describe Mill here as a utilitarian, the case can be made that he is either a pluralistic consequentialist or not a consequentialist at all.

([22]) Sidgwick, *Methods of Ethics*, 494.

([23]) G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1971; originally published 1903), 147. Should there be any doubt about his meaning, a paragraph later he emphasizes that "Our duty … can only be defined as that action, which will cause more good to exist in the Universe than any possible alternative."

([24]) Brian Barry, "And Who Is My Neighbor?," *Yale Law Journal* 88 (1979), 639n37 (a review essay of Charles Fried, *Right and Wrong*). I discuss this matter further in Judith Lichtenberg, "The Right, the All Right, and the Good," *Yale Law Journal* 92 (1983) (a review essay of Samuel Scheffler, *The Rejection of Consequentialism*). Barry tells us that the term "consequentialism" first appeared in G. E. M. Anscombe's "Modern Moral Philosophy," *Philosophy* 33 (1958); Anscombe takes credit for it on p. 10 of the essay.

([25]) Shelly Kagan, *The Limits of Morality* (Oxford: Clarendon Press, 1989), xi.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([26]) Singer asserts that, unlike effective altruists who donate to one or two charities about which they have evidence of effectiveness, “those who give small amounts to many charities are not so interested in whether what they are doing helps others—psychologists call them warm glow givers” (*The Most Good You Can Do*, 5). So on this view those who spread their donations more broadly are not simply less effective; they act out of self-interest. Singer gives no warrant for maligning their motives. At the very least one would have to show that such givers are culpably ignorant—that they have been informed or ought to know about the “scientific findings” of EA and have ignored its message.

([27]) Walter Sinnott-Armstrong, “Consequentialism,” Stanford Encyclopedia of Philosophy, at https://plato.stanford.edu/archives/sum2019/entries/consequentialism/

([28]) Brad Hooker, “Rule Consequentialism,” *Stanford Encyclopedia of Philosophy*, at https://plato.stanford.edu/archives/win2016/entries/consequentialism-rule/. Some would describe this view as global consequentialism (see Greaves, Chapter 22, this volume) rather than consequentialism full-stop. Douglas Portmore describes himself as an act consequentialist; he believes we must judge acts in consequentialist terms but denies that whether a person is blameworthy or virtuous or whether a belief is rational depends on whether blaming the person or having certain character traits or forming the belief would maximize the good. See Portmore, *Commonsense Consequentialism: Wherein Morality Meets Rationality* (Oxford: Oxford University Press, 2011). Here I stick to the more standard and old-fashioned understandings of consequentialism.

([29]) https://80000hours.org/key-ideas/

([30]) Other things being equal, you can make more of a difference when an important issue is neglected than if it isn’t. Here I mostly focus on global health and poverty, which have been and remain centerpieces of the EA movement.

([31]) Branco Milanovic, *The Haves and the Have-Nots: A Brief and Idiosyncratic History of Global Inequality* (New York: Basic Books, 2011), 116.

([32]) https://80000hours.org/key-ideas/

([33]) MacAskill, *Doing Good Better*, 76–77.

([34]) Singer, *The Most Good You Can Do*, 3–4. Some will worry about the harms such careers might cause. In his book MacAskill describes a documentary filmmaker who criticizes one of his own subjects, a cosmetic surgeon to the stars, for wasting his talent rather than saving lives. MacAskill argues that the filmmaker’s attitude “is misplaced. It’s the cosmetic surgeon’s decision about how to spend his money that really matters” (*Doing Good Better*, 78). (According to Srinivasan, MacAskill no longer recommends that people take jobs that cause direct harm. “Stop the Robot Apocalypse.”)

([35]) For such criticisms see, e.g., the brief essays in the symposium on EA, “The Logic of Effective Altruism” (*Boston Review* 40, July/August 2015), by Daron Acemoglu, Emma

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Saunders-Hastings, Angus Deaton, Jennifer Rubenstein, Leila Janah, Larissa MacFarquhar, and others.

([36]) Peter Singer, “The Logic of EA,” *Boston Review* 40, July/August 2015, 31.

([37]) Douglas Portmore points out that “Even hedonists have to rely on intuitions. For instance, even the simplest Benthamite hedonist must rely on intuition to determine how to make a trade-off between the duration of pleasure and the intensity of pleasure” (personal correspondence). Some will take this as a defense of pluralistic consequentialism, others as a rebuttal of even simple utilitarianism.

([38]) GiveWell, “Top Charities,” at https://www.givewell.org/charities/top-charities

([39]) See Cochrane Review, “Deworming Children in Developing Countries,” https://www.cochrane.org/CD000371/INFECTN_deworming-school-children-developing-countries; Jason R. Andrews et al., “The Benefits of Mass Deworming on Health Outcomes: New Evidence Synthesis, the Debate Persists,” *Lancet* 5, no. 1, January 1, 2017, https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(16)30333-3/fulltext; Susan Brink, “The Debate Is On: To Deworm or Not to Deworm,” NPR, November 15, 2015, https://www.npr.org/sections/goatsandsoda/2015/11/15/452298843/the-debate-is-on-to-deworm-or-not-to-deworm?utm:source=facebook.com&utm:medium=social&utm:campaign=npr&utm:term=nprnews&utm:con
utm:source=facebook.com&utm:medium=social&utm:campaign=npr&utm:term=nprnews&utm:con

([40]) Deaton, “The Logic of EA,” *Boston Review* 40, July/August 2015, 21.

([41]) James Lenman, “Consequentialism and Cluelessness,” *Philosophy & Public Affairs* 29 (2000), 348.

([42]) MacAskill, *Doing Good Better*, 41–42. See also Wise, “No One Is a Statistic,” *Giving Gladly*, October 10, 2018. Wise describes “a mother of seven who lives in rural Guatemala and has cervical cancer. The doctor treating her knows that screening other women for cancer is more cost-effective than treating this woman, and that the community doesn’t have enough money to fully fund both.” These other women might be considered “mere statistics,” and treating the one who already has cancer might seem more humane. Wise wisely expresses the agonizing dilemma: “Here’s the thing about those ‘statistics’: they’re all individuals.”

([43]) Singer, *The Most Good You Can Do*, 89.

([44]) MacFarquhar, “The Logic of Effective Altruism,” 25. For an account of altruists excruciatingly in touch with their moral roots, see MacFarquhar, *Strangers Drowning.*

([45]) Jeske, Chapter 12, this volume.

([46]) John Cottingham, “Partialism, Favouritism, and Morality,” *Philosophical Quarterly* 36 (1986), 357–358; quoted in Jeske, Chapter 12, this volume.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([47]) Even the weaker claim that one is permitted to be partial is expressed in deontic terms. But see Bernard Williams for the "one thought too many" argument ("Persons, Character, and Morality," in Williams, *Moral Luck* [Cambridge: Cambridge University Press, 1981], 18). Williams famously asserts that if a person had to choose whether to save his wife or a stranger, it would be "one thought too many" for him to think "It is morally permissible for me to save my wife." Unlike the standard partialist critique and the one I make here, in such situations Williams objects to thinking in moral terms at all.

([48]) As I suggested in note 28, some new and sophisticated versions of consequentialism try to avoid this.

([49]) I leave aside here the problem of comparing absolute and relative poverty and its import: whether even if, e.g., poor Americans have more to eat, their deprivations relative to others in their society is a further element that must be considered in evaluating their absolute level of well-being. See Lichtenberg, *Distant Strangers*, chaps. 5 and 6.

([50]) Philippa Foot, "Utilitarianism and the Virtues," Presidential Address to the 57th Annual Meeting of the Pacific Division of the American Philosophical Association (*Proceedings and Addresses of the APA* 57 [1983], 273. For the collapse argument see David Lyons, *Forms and Limits of Utilitarianism* (Oxford: Oxford University Press, 1967); and the discussion in Hooker, Chapter 23, this volume.

([51]) There are other important distinctions bearing on these questions that I cannot pursue here. One is between consequentialism as a criterion of rightness versus consequentialism as a decision procedure (see Peter Railton, "Alienation, Consequentialism, and the Demands of Morality," *Philosophy & Public Affairs* 13 [1984]). Another is between how I should go about deciding what I myself should do versus what acts or procedures I recommend to others (a significant distinction especially for writers, teachers, and public intellectuals). Related is the distinction between a public and an esoteric morality ("a Utilitarian may reasonably desire, on Utilitarian principles, that some of his conclusions should be rejected by mankind generally"; Sidgwick, *Methods of Ethics*, 490).

([52]) See Sidgwick, *Methods of Ethics.* Sidgwick recognizes that such conclusions are "of a paradoxical character" and that "the moral consciousness of a plain man broadly repudiates the general notion of an esoteric morality" (489–490).

([53]) This paper draws in a few places from Judith Lichtenberg, "Peter Singer's Extremely Altruistic Heirs," *The New Republic*, November 30, 2015, and Judith Lichtenberg, "The Right, the All Right, and the Good." I am grateful to Douglas Portmore and Victor Kumar for helpful comments on an earlier draft.

Judith Lichtenberg

Judith Lichtenberg is Professor Emerita of Philosophy at Georgetown University. Her primary fields of interest are international and domestic justice, moral psychology, nationalism, war, and higher education. Her book Distant Strangers: Ethics, Psychol-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ogy, and Global Poverty was published by Cambridge University Press in 2014. With Robert Fullinwider, she coauthored Leveling the Playing Field: Justice, Politics, and College Admissions (2004); she is the editor of Democracy and the Mass Media (1990). For the last several years she has been teaching philosophy at Jessup Correctional Institution in Maryland and at the D.C. Jail in Washington.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Act Consequentialism and the No-Difference Challenge

Holly Lawford-Smith and William Tuckwell

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.29

Abstract and Keywords

According to act-consequentialism, only actions that make a difference to an outcome can be morally bad. Yet, there are classes of actions that don't make a difference, but nevertheless seem to be morally bad. Explaining how such non-difference making actions are morally bad presents a challenge for act-consequentialism: the no-difference challenge. In this chapter we go into detail on what the no-difference challenge is, focusing in particular on act consequentialism. We talk about how different theories of causation affect the no-difference challenge; how the challenge shows up in real-world cases, including voting, global labor injustice, global poverty, and climate change; and we work through a number of the solutions to the challenge that have been offered, arguing that many fail to actually meet it. We defend and extend one solution that does, and we present a further solution of our own.

Keywords: act consequentialism, no-difference challenge, difference-making, thresholds, expectations, insignificance, NESS conditions, Parfit

1. Introduction

IN their 1975 paper "It Makes No Difference Whether or Not I Do It," Jonathan Glover and Martin Scott-Taggart introduced the no-difference challenge to philosophers. In their setup, they modified a well-known case from Bernard Williams in his exchange with Jack Smart (Smart and Williams 1973). In Williams's case, George the chemist faced the decision of taking a job in a laboratory researching chemical and biological warfare, and he was reluctant to accept because of his reservations about the work. George's taking the job was likely to make the world a little better, because "if George refuses the job, it will certainly go to a contemporary of George's who is not inhibited by any such scruples and is likely if appointed to push along the research with greater zeal than George would" (Smart and Williams 1973, 98).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

In Glover's retelling, the effect of a scientist's taking the job is *neutral*: she agrees that it would be better if her country did not sponsor research into chemical and biological warfare, but she reasons, correctly, "[i]f I don't do it, someone else will" (Glover and Scott-Taggart 1975, 171). There are many real-life cases with this structure, like the jobs of hired assassins, controllers of gas supplies in concentration camps, police torturers, and so on (Glover and Scott-Taggart 1975, 171). There are two ways that the effects of a person's actions can be neutral, the first by making no difference, and the second by making an insignificant difference. We'll discuss both as part of the no-difference challenge.

(p. 635) The no-difference challenge is internal to specific moral theories that are committed to the rightness or wrongness of actions (or inactions) depending on the difference that they make, for example to overall utility, or to making a person worse off than she otherwise would have been. Paradigmatically, this is a challenge within *act consequentialism*. As Frank Jackson states it: "our actions make a difference. [… T]he morality of an action depends on the difference it makes; it depends, that is, on the relationship between what would be the case were the act performed and what would be the case were the act not performed" (Jackson 1987, 94).

It is an objection to act consequentialist views (but also other difference-making views) that there are classes of actions that don't make a difference *and yet* we seem to have a strong intuition that those actions should not be performed. Our intuitions suggest that these actions are wrong; the argument from no difference (including insignificant difference) is that they're not wrong *because* they make no difference. Cases where the actions of many different people add up to cause harm at the level of the collective are prominent examples of where the no-difference challenge arises, which we will say more about later in the chapter.

Some people are happy to concede the point and agree that actions that make no difference can't be wrong. These people simply bite the bullet on the intuitive moral wrongness of George taking the job; our intuitions do not always track the correct moral theory. Others take up the *no-difference challenge.* To meet it, one must either explain how an action that appears to make no difference in fact does, or one must provide an account of why a person should not perform an action, *even though* it makes no difference. Note that the no-difference challenge cannot be met by changing the moral framework. There are moral frameworks in which the argument from no-difference does not have bite, because those frameworks do not operate in terms of difference-making. For example, virtue theory is concerned primarily with the quality of a person's character, and only secondarily with the actions that flow from that character. So a virtue theory will likely be able to deliver the verdict that the scientist should not take the job researching chemical and biological warfare, even if her doing so would make the world neither better nor worse. Indeed, such a theory might even be able to deliver this verdict in the case of George the chemist, whose taking the job would make the world better. But changing the moral framework does not *meet* the no-difference challenge head-on and provide an answer; it *avoids* the no-difference challenge by moving to a framework in which it does not arise. Thus a theorist who wants to meet the no-difference challenge by taking the second strat-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

egy—agreeing that an action makes no difference but arguing that it is wrong anyway—must do this *within the constraints* of a theory still plausibly described as act consequentialism.

While we focus in this chapter on the challenge to act consequentialism, rather than to all theories that incorporate a difference-making element, what we say will generalize with the relevant modifications for those theories' further commitments. In section 2 we'll talk about causation, so as to get a better handle on how to understand the idea of difference-making. In Section 3 we'll talk about difference-making problems in moral philosophy, leaving out the specific application to climate ethics, which we'll (p. 636) address separately in section 4. In section 5 we'll canvass some of the solutions to the difference-making challenge that have been offered across the literature and put forward a couple of our own.

2. Causation

If the morality of an action depends on the difference it makes, then the first thing we need to get a handle on is what exactly it means for an action to make a difference. Did your action make a difference to the bank heist when it was *putting stolen money into bags and leaving with them?* How about when it was *keeping lookout while others stole the money?* And how about when it was *providing the security card to get into the bank*, which the bank heist could not have gone ahead without?

Theorists of causation have been interested in difference-making, as an answer to the question of what it means to *cause* something. Many investigations into causation take the following natural thought about causation as a point of departure:

> We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it.(Lewis 1986, 160–161)

Carolina Sartorio labels this *the difference-making idea* (Sartorio 2005, 71). Sartorio takes this idea to be so plausible that she argues that any acceptable account of causation must respect it; accounts of causation should be constrained by the difference-making idea.

In this section, we'll run through three prominent accounts of causation: causation as production, causation as counterfactual dependence, and causation as influence. We'll detail the differences in the understanding of difference-making that result. We'll also introduce two complications, the first to do with another's action *pre-empting* yours, the second to do with your action *overdetermining* harm. Arguably, neither pre-empting nor overdetermining contributions are difference-makers, and if they're not difference-makers, they're not causes either (at least if we agree with Sartorio).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

First up, *productive theories of causation*. Productive theories take causation to consist in the transmission of energy from cause to effect. One prominent example of a productive theory is Phil Dowe's (2000). According to Dowe, to say that *c* is a cause of *e* is to say that there is a transfer of conserved quantity—force or energy—from *c* to *e*. For example, on this view we'd say that Suzy's rock throw caused the window to smash as a result of transferring energy from the thrown rock to the window.

One objection to productive theories is that they cannot account for causation by omission. For example, we tend to have the intuition that not watering the plant can cause the plant to die, but there is no transfer of energy between my *not watering* and the plant dying. My not watering might be said to make a difference, and so to be a cause; if that is right, then the productive theory of causation is wrong.

(p. 637) Next, *counterfactual theories of causation.* According to counterfactual theories, *c* is a cause of *e* iff had *c* not occurred, then *e* would not have occurred either. For example, the counterfactual theory says that Suzy's rock throw is the cause of the window being smashed because if Suzy had not thrown the rock, then the window would not have shattered. David Lewis is the most famous advocate of this view (Lewis 1973).

Simple glosses on counterfactual theories are subject to some easy counterexamples that show they are unable to respect the difference-making idea. This is true of *pre-emption* cases: suppose that Suzy throws a rock at a window and, as a result, the window smashes. Suppose further that Billy is waiting in reserve. Had Suzy not thrown her rock, Billy would have, and the window still would have shattered. Intuitively, Suzy's rock throw made a difference to, and caused, the window to shatter. But, on the simple reading, the counterfactual theory reaches the verdict that Suzy did not cause—or make a difference to—the window smashing. This is because even if Suzy had not thrown her rock, Billy would have thrown his, and the window smashing would still have occurred. The window's smashing does not depend on Suzy.

A further class of counterexamples that have proved challenging for simple glosses on counterfactual theories are *overdetermination* cases: suppose that Suzy and Billy both throw their rocks at the window at the same time. Suppose that both rocks hit the window at the same time, and that the window smashes. Either of Suzy or Billy's rock throws alone would have been sufficient to smash the window. If Suzy did not throw her rock, the window still would have smashed, in roughly the same way, as a result of Billy's throw. If Billy did not throw his rock, the window still would still have smashed, in roughly the same way, as a result of Suzy's throw. So the verdict of the counterfactual theory is that neither Suzy's nor Billy's rock throw caused—or made a difference to—the window smashing. But this is surely the wrong result; the window being smashed was caused by *something*, and Suzy and/or Billy's rock throw seem to be that thing. (For the argument that there are no overdetermination cases, see Bunzl 1979.) A more sophisticated reading of the counterfactual theory, in terms of stepwise chains of counterfactual dependence, can arguably avoid these problems (Lewis 1973; see discussion in Eriksson unpublished manuscript).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The final account of causation that we'll mention is Lewis's (2000, 2004) *influence theory of causation*. The influence theory says that causation is a matter of counterfactual co-variation between modally fragile alternatives of two events: a cause, *c*, and an effect, *e*. Consider the following example: suppose that Suzy throws her rock at a particular angle. Call this event *c*. And suppose that Suzy throwing her rock as she does means that the window smashes in some particular way—the glass splinters in some particular direction. Call this event *e*. A modally fragile alternative of *c* is a slight variation in the way that Suzy throws the rock—she throws the rock at a slightly different angle. A modally fragile alternative of *e* is a slight variation of the way in which the window smashes—the glass splinters in a different direction. According to the influence theory, if there is counterfactual covariation between these modally fragile alternatives of *c* and *e*, then there is causal influence of *c* on *e* (see Bernstein forthcoming for further discussion of (p. 638) Lewis's various theories of causation). On this theory, difference-making is possible even when there appears to be pre-emption or overdetermination.

To finish up our discussion of causation, it is worth mentioning a problem that arises when considering the relationship between causation and moral responsibility that Sara Bernstein (2017) has drawn attention to. Bernstein claims that the following is a generally accepted principle of moral responsibility:

> *Proportionality:* An agent's moral responsibility for an outcome is proportionate to her actual causal contribution. (Bernstein 2017, 167)

According to *Proportionality*, the greater an agent's causal contribution to an outcome, the greater her moral responsibility for the outcome (ceteris paribus). However, Bernstein argues that no leading theory of causation can capture the idea that differences in causal contribution make a proportionate difference to moral responsibility (Bernstein 2017, 166). She illustrates this with the following case:

> *Bad Politician*: Bad Scientist hoists a weapon onto a launch pad for mere testing. But Bad Politician gains access to the weapon's computerized control system, and presses the launch button, launching the weapon. (Bernstein 2017, 174)

Intuitively, Bad Politician is more morally responsible for the launch of the weapon than is Bad Scientist. But none of the theories of causation that we've considered can account for this. According to the productive theory of causation, given that Bad Scientist transfers more conserved energy to the weapon than does Bad Politician, Bad Scientist is more morally responsible (Bernstein 2017, 174). The counterfactual theory of causation cannot account for Bad Politician's being more of a cause, and so more morally responsible, for the launch of the weapon because counterfactual dependence either holds or it does not; something either is or it is not a cause. Something cannot be more or less of a cause than something else (Bernstein 2017, 175). Finally, the influence theory of causation cannot account for the difference in moral responsibility either. To see this, compare what would happen if there were a variation in Bad Politician's button pushing—he pushed it at a slightly earlier time—with what would happen if there were a slight variation in Bad Scientists placement of the weapon of the launch pad—he put it on in a slightly wonky fash-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ion. In the former case the weapon would land at a slightly earlier time, in the latter the weapon would land in a different location where it causes less harm. The pattern of counterfactual covariation is stronger with alterations in Bad Scientist's action than it is with Bad Politician's actions, and so the actions of Bad Scientist are more of cause than the actions of Bad Politician. From the perspective of carrying out a moral assessment, this is intuitively the wrong result (Bernstein 2017, 176). One obvious move here is to deny that moral responsibility is proportionate to an agent's causal contribution to an outcome. An alternative is to think that causal contribution is a necessary condition of moral responsibility, but that degrees of moral responsibility are established in other ways (this is a particularly appealing alternative when thinking (p. 639) about moral responsibility for outcomes that are collectively caused; see discussion in Lawford-Smith 2019, chap. 6).

(For further discussion of causation relevant to understanding the no-difference challenge, see also Braham and van Hees 2009; Fenton-Glynn 2013; Hitchcock 1993; Kaiserman 2016; Kment 2010; Menzies 2004; Nye 2009; Sartorio 2005; Schaffer 2003; Waters 2007; Weslake forthcoming; and Williamson 2010.)

3. The No-Difference Challenge in Moral Philosophy

The no-difference challenge is discussed using many different types of cases: stealing beans from the hungry (Glover and Scott-Taggart 1975; Jackson 1987); giving water to dehydrated people (Parfit 1984, 76–78); turning a torture dial up (Arntzenius and McCarthy 1997; Quinn 1990); polluting the air, polluting the water supply, buying factory-farmed meat (Kagan 2011); contributing to climate change through driving cars (Sinnott-Armstrong 2005). But these are mostly superficial, in use only to make the no-difference challenge vivid rather than to apply the challenge to a specific case. The four main real-world areas in which it more systematically applies are voting, global labor injustice, global poverty, and climate change. We'll talk about the first three here and address the fourth separately in section 4.

Before we get into the cases, a clarification is needed. The no-difference challenge is about actions not making a difference to specific outcomes, like the result of an election, or climate change harms. It's not met by showing that actions make a difference to *something other than those specific outcomes.* So, for example, attempts to resolve the no-difference challenge in the case of voting by stipulating that you've promised a friend to vote, or that voting expresses support for democracy, do not meet the challenge (see also discussion of this clarification in Nefsky 2019, 2).

Voting. The first place where the difference-making challenge shows up is in the question of whether an individual should vote. In particular, theorists ask whether it is rational to vote, and whether there is a moral duty to vote (Brennan 2016). Whether it is rational to vote depends on how likely it is that the individual's vote will make a difference (Brennan 2016). This is the likelihood of being a "tie-breaker" in an electoral outcome, namely, that

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

the election comes down to a single vote, and that the individual's vote is the deciding vote. Some have put the best chance of this at 1 in 10 million, for an individual living in a swing state in America and voting for a major-party candidate (Edlin, Gelman, and Kaplan 2007; discussed in Brennan 2016).

This raises the question of whether it's enough to defeat the no-difference challenge to establish *some expectation of making a difference*; after all, a 1 in 10 million chance of making a difference is not *no* chance of making a difference. There might be further things that an act consequentialist would want to say about *which* of the difference-making (p. 640) actions you should choose, in particular that you should choose the one with the highest utility. It is unlikely that an act with such a small chance of making a difference will turn out to be the act, from all of those available to the individual, with the highest utility. But it is not impossible.

Whether one has a moral duty to vote might be thought to depend on similar sorts of considerations, for example that we ought to vote because we have a duty to protect ourselves, or help others, or produce a good government (Brennan 2016). But these are all themselves vulnerable to no-difference challenge. Some theorists have turned to alternative explanations of why we ought to vote, instead, such as that voting has expressive or symbolic value (Brennan and Lomasky 1993), or that voting avoids complicity (Beerbohm 2012). But notice that these do not meet the difference-making challenge, they rather shift to a moral framework in which it does not arise.

Global labor injustice. The discussion of difference-making in the context of global labor injustice is best-known from Shelly Kagan's (2011) paper "Do I Make a Difference?" Kagan introduces the cases of consumer purchases of chicken meat. He worries that consequentialism cannot condemn our purchases of dead chicken carcasses at the butcher's counter, because "whether or not I buy a chicken makes no difference at all to how many chickens are ordered by the store—and thus no difference in the lives of any chickens" (Kagan 2011, 110).

The chicken case is a good example of a collective action problem, because most of our purchases taken alone make no difference to the harms of the poultry industry, but *all* of our purchases taken together make a huge difference ("if several hundred thousand fewer chickens were sold this week, the chicken industry would dramatically reduce the number of chickens it tortures" (Kagan 2011, 111). In standard collective action cases, what is in the best interests of a group of people stands in tension with what is in the best interests of any individual member of the group; for example, it is in my interests to overfish because I get more fish to eat or sell, but if we *all* overfish, then the fishery collapses and we all get nothing. But the term has also been extended to cases where there are simply different effects or incentives at the level of the group and the level of the individual.

What applies to the chicken is even more complicated in the case of sweatshop labor, which produces a range of different products that we buy. For example, consider a simple t-shirt. In buying a t-shirt from a shop, we create demand for t-shirts. We usually don't

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

know whether our purchase was the one that caused a bunch more t-shirts to be ordered from the supplier. There is also some further uncertainty as a result of interim actors in the supply chain—perhaps the shop you buy from doesn't reorder directly from the manufacturer, but from some further suppliers, who in turn don't reorder until they hit a certain number of requests (see further discussion in Eriksson unpublished manuscript). In addition, there's a question of what it is I cause *when* I cause a reordering. It might be "the violation of the labor rights of a particular worker in a particular location at a particular time" or her falling below a particular threshold level of well-being (Lawford-Smith 2018).

But even this will depend on a number of further things, such as exactly what the production processes of the sweatshop are (if multiple workers produce parts of one good, (p. 641) then you can at most be a partial cause of harm to a group, whereas if each worker produces a complete good, then you can be a full cause of harm to that worker), and whether there are some number of hours that the worker can work without injustice (which would be true if the labor injustice consists *only* in her long hours, as opposed to the working conditions themselves) (see further discussion in Lawford-Smith 2018). There's also a further complication here in terms of what can happen if a reorder isn't caused, in terms of making sweatshop workers *even* worse off than they already were.

Global poverty. Finally, global poverty is a third area in which difference-making challenges arise, most familiarly through the idea that nothing I can do will make a difference to poverty, disease, homelessness, child marriage, and so on. It matters here how the outcomes are described: should my donation of $5 to the Against Malaria Foundation make a difference to *global poverty*, holistically described? Or should it only make a difference to the suffering of one individual, that is, the person who, because she has a malaria net, may thereby be prevented from contracting malaria? We get a different result for thinking about difference-making depending on the level of description. Effective altruists like Peter Singer (2015) and Will MacAskill (2015) have been careful to meet the difference-making objection by pointing to interventions where small donations or contributions in other metrics *do* make a positive difference.

(For further discussion of the no-difference challenge in moral philosophy excluding climate ethics, see also Barnett 2018; Barry and Øverland 2016; Bernstein 2017; Budolfson forthcoming; Cullity 2000; McGrath 2003; Nefsky 2012; 2017; Norcross 2004; Pinkert 2015; Sartorio 2007; 2010; and Talbot 2018.)

4. The No-Difference Challenge in Climate Ethics

Climate change has a similar structure to the cases discussed in section 3, namely that it is a large-scale collective action problem. People from all over the world emit various greenhouse gases (GHGs). It is unlikely (although not impossible) that any one emitting action causes harm in isolation, and yet many such actions come together to cause harm.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

But climate change has three further features which make it more intractable than the problems discussed already.

First, greenhouse gases feed into *one central system*, and this system produces harm. This is unlike global labor injustice or global poverty, which involve many different systems (although it is like voting, considered within one particular democratic country). That means we're always balancing the chance of being a difference-maker to climate harms against the chance of *all* other contributions from the approximately 7.7 billion people in the world (as of April 2019) being difference-makers.

Second, the effect of an individual's emissions is not only hard to know but impossible in principle to know. Because the system is central, and because there are emissions of (p. 642) different types being made at different times and in different quantities, we can only know in the most abstract terms what we *might* be contributing to. Scientists can give more concrete predictions about macro-level harms, like rainforest death or permafrost melt, but it's unlikely *your* actions will be a difference-maker in those. But they can't give concrete predictions about micro-level harms—for example, that there will be a particular extreme weather event and it will be *this* much worse than it could have been and hurt *this* many more people than it would have without your contribution—when it's more likely that your actions *will* be a difference-maker in those. In the labor injustice case you could actually find out what the reordering threshold is and how many purchases had been made already; in the climate change case you can't. So it's harder for you to make decisions about what it's reasonable to do.

Third, there's a huge lag between your actions and the effects of your actions. GHGs stay around in the atmosphere, a proportion of them for a very long time, doing harm. In the case of causing particular of the harms of global poverty, or causing the alleviation of particular of those harms, there might be a delay of some months or even years between you making a donation and a project being completed. But in the case of emitting carbon and so causing particular of the harms of climate change, or reducing your emissions so as to mitigate some of those harms, the effects may take fifty or a hundred years to be seen (and even then, because of the second problem, you won't know about them).

The approach in climate ethics for those who take the no-difference challenge seriously has generally been to try to argue that indeed single acts associated with emissions *do* make a difference. The best-known defense of this claim comes from John Broome. He argues that individual emissions (the emissions associated with a single individual's specific emitting action) cause harm—"every bit of emission that you do cause is harmful" (Broome 2012, 77)—on the basis that emissions spend a long time in the atmosphere and therefore have "innumerable opportunities to cause harm" (discussed in Lawford-Smith 2016a, 131). Broome concedes that it's not guaranteed that a particular emission will do harm, but if we considered a number of your emissions, it's extremely likely that one or more of them would do.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

According to Broome, we have a duty not to cause harm, so we have a duty not to emit (note that one need not be an act consequentialist to accept this claim). If he's right about individual emissions doing harm, then he's right that on the act consequentialist view, and other views that care about difference-making, agents have a reason not to emit.

Note one idiosyncrasy in Broome's view, which is that he endorses Bernard Williams's reasoning against Jack Smart in their exchange over utilitarianism, and so maintains that "the injustice [of emitting] consists in harming, not in merely causing more harm to be done" (Broome 2012, 84; see also discussion in Lawford-Smith 2016a, 132). By this he means it can matter to the overall utility calculus that *you* did harm, through your actions. This suggests that he would reject the principle we started with, that the morality of your (emissions) action depends on the difference that it makes. So in the cases where your action is a cause of harm but does not make the world any better or worse **(p. 643)** (say, because if you hadn't emitted someone else would have), Broome's view of emissions as harms would not offer a solution to the no-difference challenge. Fortunately, these coincidental neutrality cases are likely to be rare.

Another attempt to meet the no-difference challenge in climate ethics comes from Avram Hiller (2011). Hiller argues that it's not the case that individual actions are too causally insignificant to make a difference with respect to climate change, and that as such we can evaluate individual emitting actions as morally wrong. The force of Hiller's objection comes from shifting the focus from *actual* harm caused by individual emitting actions to a focus on the harm that can be *expected* to result from such actions. Hiller argues that if we accept the plausible moral principle that "it is prima facie wrong to perform an act which has an *expected* amount of harm greater than another easily available alternative" (2011, 352), and we come to see that individual actions to which there is an easily available alternative do in fact cause some not insignificant harm, we will see that individual actions are difference-making in a morally significant way.

Hiller relies upon some empirical findings to make his case: John Nolt's (2011) finding that on average an American's lifetime GHG-emitting activities cause serious harm to at least one person, generally in the developing world, and data from the National Academy of Sciences that show that a twenty-five-mile car journey approximates a quarter of a day's worth of the average American's emissions. Using this data, Hiller claims that a quarter of a day's worth of emissions causes a quarter of a day's worth of serious harm: "going on a Sunday drive is the moral equivalent of ruining someone's afternoon" (Hiller 2011, 357). On the assumption that Nolt's calculations are correct—Hiller takes these to be the best estimations that we currently have, even if some might have some methodological concerns—he concludes that, given that ruining someone's Sunday afternoon is a pretty mean-spirited thing to do, individual actions like going on a Sunday afternoon drive are prima facie wrong in a noninsignificant way.

However, unlike Broome, Hiller's reasoning relies on a "share of the total" view of expected harm. His method is to estimate the amount of GHG emitted by the one drive; estimate the total amount of GHG emissions responsible for climate change; estimate the total

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

amount of harm that climate change will cause; and then make a calculation based on these variables (Hiller 2011, 357). This looks more like *joint* causation of harm, with shares attributed back to individuals, than individual causation of harm through individual actions. It's not clear that individual emitters actually *make a difference*, even an expected difference, on Hiller's view, and so that individual emitters cause harm at all.

(For further discussion of the no-difference challenge for climate ethics, see also Almassi 2012; Andreou 2006; Barry and Øverland 2015; Garvey 2011; Gesang 2017; Gunnemyr 2019; Kingston and Sinnott-Armstrong 2018; Lawford-Smith 2016a; 2016b; Morgan-Knapp and Goodman 2015; Pellegrino forthcoming; Rendall 2015; Sandberg 2011; Sinnott-Armstrong 2005; Spiekermann 2014; Schwenkenbecher 2014; Vance 2017; and Vanderheiden 2007.)

(p. 644) 5. (Further) Solutions to the No-Difference Challenge

Many people have attempted solutions to the no-difference challenge (beyond simply biting the bullet on the actions in question not being wrong). We run through some of the most noteworthy next and propose a couple of our own.

Group wrong without individual wrong. Frank Jackson argues in his 1987 paper "Group morality" that when a number of individuals perform actions that are not wrong taken individually, but which cause harm when taken together, *the group does wrong.* He insists that it is not merely that wrong "happens"; rather, wrong is done by the group (Jackson 1987, 100). Consider two versions of a case. There are a thousand villagers each with a thousand beans, and there are a thousand bandits looking to steal the beans. In one version, the bandits steal "vertically," meaning one bandit steals all one thousand beans from one villager. In the other version, the bandits steal "horizontally," meaning every bandit steals one bean from each of the thousand villagers. Jackson's point is that "[s]witching from the vertical to the horizontal version makes not one iota of difference to the moral status of the group action; it is only the standings of the individual actions that are affected" (Jackson 1987, 101). He does not limit his claim to organized groups, so in principle he's allowing group wrongdoing in cases of entirely unconnected and uncoordinated individuals. Jackson's solution captures the intuition that *something* morally wrong has happened, but explains this as occurring at the level of the group rather than the level of the individual.

Membership in the smallest set. In "Five Mistakes in Moral Mathematics," chapter 3 of the canonical *Reasons and Persons* ([1984] 2003), Derek Parfit argues that one can be considered a contributor to harm when one is a member of the *smallest set such that if they didn't act the harm wouldn't occur*. His motivation is to avoid the problems of overdermination and pre-emption (see discussion in section 2). Parfit worries that even when we can identify a set of people such that had they acted differently I would not have been harmed, it will nonetheless be true that had they *and Fred Astaire* acted differently, I

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

would not have been harmed (Parfit [1984] 2003, 72). In other words, we can always add arbitrary members to the set and the claim will still be true. Thus we need to specify that the set is "the *smallest* group of whom it is true that, if they had all acted differently, the other people would not have been harmed, or benefited" (Parfit [1984] 2003, 72).

Beth Kahn (forthcoming) argues against this solution on the grounds that it lets other contributors off the hook; if ten factories all contribute to pollution, and that pollution does harm, it's odd to say that only two factories actually count as contributing to harm on the grounds that the two make up the smallest set such that if they didn't contribute, the pollution harms would not occur. On her view, *all* contributors should be on the hook as contributors to harm, and they have reason not to contribute in order to avoid that.

(p. 645) *Share of total benefit/minus reduction in benefit I cause by joining/plus increase in benefit I cause by joining.* Parfit set his solution up against three others. The first was the simpler idea that one's contribution to harm is equivalent to her share in the total harm done. For example, there are a hundred miners trapped in a shaft, and I could join four other people to save them. If I do, I am responsible for one-fifth of the lives saved; twenty lives (Parfit [1984] 2003, 68–69).

The second was the slightly more complicated idea that we should also factor in any reduction in benefits that I cause by joining. Imagine instead, as Parfit does, that I could join four others in rescuing the miners, *or* I could go off separately and save ten people by myself, and if I do that, the efforts of the four will still be sufficient to saving the hundred miners. If I chose to help the four anyway, then we must factor in the negative difference I make. On the one hand, I am one of five people who rescue one hundred miners, so I save twenty lives. On the other hand, had I not joined the four, they each would have saved twenty-five lives each. By my joining, I made it the case that they each saved five fewer people, for a total of twenty people. When I subtract the reduction in benefit that I cause from the benefit I contributed to, I end up at zero. This means I have a reason to choose to save the ten by myself, instead (Parfit [1984] 2003, 69). This is intuitively correct, because a greater number of people are saved overall (the one hundred miners plus the ten others).

The third solution extends the second, by saying that we should factor in not only a reduction in benefits caused by my joining, but also an increase. Imagine, as Parfit does, that if I joined three others we could together save a hundred people, and so save twenty-five lives each on the share-of-the-total view, but I could also go off separately and save fifty people by myself. However, if I did that, the hundred people would not be saved: they need me. By joining them I cause an increase in their shares, from zero to twenty-five. My individual share is twenty-five lives, but I also cause the increase in others' share, totaling seventy-five lives (Parfit [1984] 2003, 69–70). Again, this is intuitively correct, because it has me join together with others to rescue one hundred people, rather than go off by myself to rescue fifty.

As we saw already, Parfit was not satisfied with these solutions, and he defended the alternative of membership in the smallest set instead.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

NESS conditions. The idea of NESS conditions comes from Richard Wright (1985). The proposal is that what matters is whether a person's action is a "necessary element of a set of antecedent actual conditions that was sufficient for the occurrence of the consequence" (Wright 1985, quoted in Barry and Øverland 2016, 231), or more succinctly, a necessary element of a sufficient set (sufficient to producing harm). Suppose that in order for a person to die, he must be kicked one hundred times. Suppose further that he is kicked two times each by one hundred people, so that his death is twice overdetermined. We'd divide all of these kicks into all the logically possible configurations of one hundred kicks (one each by everyone; two each by half the group; four each by a quarter of the group; one by this guy, two by this guy, three by this guy… and so on) until we had all the sets on the table. Each of these sets would be sufficient for the death. Members are necessary to these sets not in any deep sense; of course, it could have been that they'd **(p. 646)** been a member of another such set instead. But once assigned, because the assignments are logically exhaustive, membership is necessary in a technical sense. Wrongful action comes cheap on this account, because it only requires membership in *a* sufficient set. So it wouldn't matter how many others were making contributions and how large their contributions were, so long as you made some contribution you'd be a member of at least one set, and thereby on the moral hook.

Deriving individual imperatives from collective imperatives. Garrett Cullity (2000) claims that the relationship between collective and individual imperatives can show us why individuals are required to act or not act in situations where individual actions make no difference, but where the aggregation of many individuals' failure to act or acting does cause a morally significant bad outcome (and why individuals are required to act in contributing to pools for beneficence). Cullity claims that in the same way that free riding involves an unfair failure to contribute to a scheme that produces a public good while the free rider at the same time benefits from that scheme, the person who fails to contribute to a pattern of action that is required if a group is to do what it ought to do collectively is failing to act in accordance with what fairness demands of them. In both cases, there is a collective imperative that ought to be met. In order for the collective imperative to be met, many individuals must make individually imperceptible contributions—either by acting, for example giving some small amount of money to poverty relief, or by refraining from acting, for example not engaging in high GHG-emitting activities. Where many others are prepared to make such contributions, the individuals who fail to act unfairly leave it up to others to do what they are not prepared to do themselves. Cullity claims these considerations of fairness allow us to derive an individual imperative from a collective one (Cullity 2000, 20–24).

Problems with these solutions. The same problem faces all the solutions presented so far, namely that it's not clear they actually meet the difference-making challenge without changing the moral framework. An act consequentialist insists that the morality of an action should be assessed as a matter of the difference it is likely to make. That is the difference *it* is likely to make, not the difference *the group* is likely to make (cf. Jackson 1987), or the difference the *smallest set* is likely to make (cf. Parfit [1984] 2003), or the difference a *sufficient set* is likely to make (cf. Wright 1985), or what it is *fair* to ask of an indi-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

vidual in light of what a collective must do (cf. Cullity 2000). If we shift focus away from individual actions and onto things like groups or sets of actions, we change the moral framework. If we don't, it's not clear why these should be considered solutions to the no-difference challenge at all.

Expectation of difference-making. Shelley Kagan explores two solutions to the no-difference challenge in his 2011 paper "Do I Make a Difference?" The first is that our actions have some chance of making—in the "causation as production" sense—a major difference (this is to draw out Parfit's point in *Reasons and Persons* that it is one of the five mistakes in moral mathematics to ignore small chances). For example, the chance of my carbon emissions being the ones to finally cross the threshold for a catastrophic climate harm occurring is very small, but if they *were* the emissions to cross the threshold, that (p. 647) would be very bad indeed. Thus I can still have a reason not to take the risk, because I cannot be sure that my action won't make a terrible negative difference.

Kagan assumes that the badness of an outcome in which it's my action that causes the harm trades off favorably against the goodness of all the outcomes in which it's *not* my action, and that the utilities will always work out this way: I may have smaller chances in some cases, but in those cases the harms will be bigger, so the upshot (that I shouldn't perform the action) will be the same (Kagan 2011, 117–121; see also discussion in Lawford-Smith 2016b).

Julia Nefsky has responded to Kagan's first solution, arguing that he makes a mistake in assuming that the utilities will always work out in such a way that the badness of the harm counterbalances the small chance of bringing it about (Nefsky 2012). Her conclusion is that "we cannot say that in every case of collective harm, each act might make a difference" (Nefsky 2012, 395). Her own preferred solution is to reject the assumption that if you don't make a difference to an outcome then you don't help to bring it about, and to shift the moral focus onto helping to bring outcomes about (Nefsky 2012, 395). But this solution will be guilty of the same problem as some of the others, namely that it sidesteps the no-difference challenge rather than answering it (which is not to say that this is the wrong move, only that it is not a move you can make while still playing the game).

In Kagan's second solution, he discusses a case in which *everyone* who contributed up to the threshold—at least, supposing that certain other conditions are met—causes the threshold to be crossed. The conditions matter: it must be that the harm is not under- or overdetermined. If it is underdetermined, so that contributions fall short of the threshold, then no one causes harm. If it is overdetermined, so that contributions surpass the threshold, then no one causes harm. Note that we've now shifted to the "causation as counterfactual dependence" sense. Only when there is exactly the number of contributions necessary to cross the threshold and no more, does everyone cause the harm, because *the harm depends counterfactually on everyone.* Let's call these groups of contributors "cohorts." What matters in making a decision about whether to act is the expectation of my action being part of such a cohort. My action makes a difference when it is, and it doesn't when it isn't.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Kagan's first solution has an interesting application in the case of climate change, where it's plausible to think there are huge numbers of micro-level thresholds (Lawford-Smith 2016b). Should we think that almost everyone is a cause of harm, because someone who is an underdeterminer or overdeterminer relative to one cohort will be someone on whom the harm is counterfactually dependent relative to another cohort? Or should we think that *because* there are multiple cohorts, *everyone* is an underdeterminer and/or overdeterminer of harm, relative to all cohorts?

If it's the latter, then Kagan's solution is no solution at all to the large-scale cases we discussed in section 3. It would only apply in the extremely limited set of cases in which the harm has not yet occurred, and it is perfectly possible that exactly as many people and no more end up contributing such that the threshold is met but not crossed. And it seems plausible that it *is* the latter. We know in the case of climate change, at least, that **(p. 648)** the type of emissions (methane, carbon, hydroflurocarbon, etc.), the timing of emissions (sooner rather than later), where the emissions occur (e.g., higher in the atmosphere vs. lower), and the amount of emissions all matter. Suppose there is a micro-level climate threshold that depends on 500,000 people in New Zealand emitting carbon by driving their cars in the month of June 2019. Suppose further that Tāne drives his car once during June 2019. Why think his action is part of the cohort that causes that harm, rather than—given the millions of people driving cars in New Zealand—that it underdetermines the harm (for all we know, insufficiently many people drove their cars within the time frame, so that particular threshold was not crossed), or that it overdetermines the harm (for all we know, many more people than necessary to the threshold's being crossed drove their cars within the time frame)? There are always more thresholds and more cohorts, so it's not clear why this *solves* the problem rather than *exacerbates* it.

A further problem with the expectation of difference-making is that it's not clear in the climate case whether the expectation is of making a *positive* or a *negative* difference. Our emissions might cause an extreme weather event to happen in a slightly different way, but for all we know that way might be worse for those whose interests we care about, not better. This issue does not affect the voting, global labor injustice, or global poverty applications, but it does affect the climate change application.

Probability of membership in actual set causing harm. Christian Barry and Gerhard Øverland argue in their (2016) book for a modification to the NESS conditions solution that solves the problem of overdetermination (which, as we have seen, is a challenge to difference-making). Instead of looking for an action's being a necessary element of a sufficient set—any one such set—they suggest we look for the *probability* of an action's being an element of *the actual* set that causes the harm (Barry and Øverland 2016, chap. 11). That is to put the solution more in terms of expectations, in the way that Kagan's (2011) solution does. Suppose that it takes fifty actions to cause a harm, so fifty is the threshold. Then the more actions there are over the threshold (e.g., there is a total of five hundred actions), the lower the probability your action has of being a contributor to the harm; and the fewer actions there are over the threshold (e.g., there is a total of fifty-one actions), the higher the probability your action has of being a contributor to harm. Because many

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

of the large-scale cases in which harm is overdetermined are such that it's epistemically opaque to us how our action relates to the harm, this is a doxastically rational way to think about difference-making. If our moral reasons track expectation of difference-making, then the higher the probability our action has of being a contributor to harm, the stronger the reason we have not to perform it (taking into account also how bad the harm is and what other benefits our action might bring).

Positive formulation of permissibility. Another solution to the no-difference challenge to act consequentialism, which we find attractive, is to interpret the act consequentialist claim *positively* rather than *negatively*. The negative interpretation is something like "it's permissible to φ so long as your φ-ing doesn't make the world worse than it otherwise would have been." The positive interpretation is something like "it's permissible to φ only if your φ-ing would make the world better than it otherwise would have been." Your taking the job in chemical or biological research, or becoming an assassin, or controlling (p. 649) the gas chambers at a concentration camp, or becoming a police torturer, all might not make the world worse, so long as there are plenty of others willing to do the job if you don't. On the negative interpretation of the act consequentialist claim, that would make it permissible for you to take those jobs, because doing so wouldn't make the world worse.

But on the positive interpretation of the act consequentialist claim, it would be impermissible for you to take those jobs *unless*—like George the chemist in Williams's version of the case—your doing so would make the world *better*. The same is true for decisions in collective action cases, such as about your personal carbon and methane emissions, or consumption decisions, or contributions to the relief of global poverty. If maintaining your emissions levels or consumption choices would make the world better (e.g., because if you don't, someone worse will step in to fill the gap), then you have a reason to do it. If it wouldn't, you don't, *even if* maintenance wouldn't make the world worse. This positive interpretation is consistent with what Parfit says in "Five Mistakes in Moral Mathematics," where he says "on any plausible moral theory, we should sometimes try to do what would benefit people most" (Parfit [1984] 2003, 70).

Forward-looking NESS conditions. As we mentioned in sections 3 and 4, the no-difference challenge has both a backward-looking application and a forward-looking application. In the former, it's about attributing wrongdoing to actions that didn't make any individual difference. In the latter, it's about giving people a reason not to perform actions, now or into the future, that they reasonably believe won't make any individual difference. This final solution is relevant only to the latter. It builds upon Barry and Øverland's (2016) proposal that what we should be tracking is the probability that my action is an element of the actual set of actions that cause the harm. If it takes one hundred kicks to kill a person, and I'm an enthusiastic person-kicker, what I'd want to know is the probability of my kick being one of the one hundred that actually causes the death.

It also builds upon the positive interpretation of act consequentialism just outlined, namely that an action is permissible only if it makes the world better than it otherwise would

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

have been. Formulating the Barry and Øverland proposal positively, what we should be tracking is the probability that my action *will be* an element of the actual set of actions that *mitigate/avoid/reduce* a harm. For example, imagine we're thinking about funding effective sewerage systems in a poor city that lacks them. If no one does anything, all of the harms that come from a lack of adequate sanitation will continue to affect the city's inhabitants. You're considering whether to give some money to the project.

The first thing to think through is the NESS conditions move: the logically complete set of constellations of actions, each of which would be sufficient to fund the sewerage project. Because there's so much poverty in the world, there are a lot of people for whom they're not in a position to contribute to this project. So the number of sufficient sets will be much smaller than for some other collective action that everyone in principle could contribute to (like vote a certain way in a global referendum, for example). The next thing to check is *how many* sets your action is an element of. Because the future is open, we don't know which will be the actual set to fund the sewerage project. This is not quite (p. 650) *probability of being in the actual set*, in the way Barry and Øverland imagined it, because there is no actual set yet. But it's a heuristic for thinking about the relevance of your contribution.

The more sets your action is an element of, the stronger a reason you have for making a contribution. In some kinds of cases, like radically reducing carbon and methane emissions in order to mitigate climate change, the actions of people in wealthy, industrialized countries might be in many if not most of the possible sets of actions sufficient to mitigation. So there is a prospective sense in which these actions are difference-makers: they can be expected to be difference-makers in avoiding harm (in making the world better), and so those who could perform them have good reasons to do so.

6. Conclusion

The no-difference challenge arises in cases where an individual's action makes no difference or makes an insignificant difference. The no-difference challenge can also appear to arise—although ultimately does not—in cases where an individual's action is unlikely to make a difference. The challenge shows up particularly in collective action cases, and it has applications in a wide range of real-life cases, including voting, global poverty, global labor injustice, and climate change. Although many have attempted to meet the challenge, solutions often sidestep the problem by changing either the subject of what a difference is made *to* or changing the moral framework (away from act utilitarianism).

We have argued that appealing solutions include Barry and Overland's (2016) idea that an action makes more or less of a difference depending on the probability of its being a member of the actual set of actions that cause a harmful outcome; and understanding the act utilitarian requirement in a positive, rather than a negative, way, so that an action is permissible only when it makes the world *better* than it otherwise would have been,

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

rather than impermissible only when it makes the world worse (which leaves neutral actions on the table).

References

Almassi, B. 2012. “Climate Change and the Ethics of Individual Emissions: A Response to Sinnott-Armstrong.” *Perspectives: International Postgraduate Journal of Philosophy* 4, no. 1: 4–21.

Andreou, C. 2006. “Environmental Damage and the Puzzle of the Self-Torturer.” *Philosophy and Public Affairs* 34, no. 1: 95–108.

Arntzenius, F., and McCarthy, D. 1997. “Self-Torture and Group Beneficence.” *Erkenntnis* 47: 129–144.

Barnett, Z. 2018. “No Free Lunch: The Significance of Tiny Contributions.” *Analysis* 78, no. 1: 3–13.

Barry, C., and Øverland, G. 2015. “Individual Responsibility for Carbon Emissions: Is There Anything Wrong With Overdetermining Harm?” In *Climate Change and Justice*, edited by J. Moss, 65–83. Cambridge: Cambridge University Press.

(p. 651) Barry, C., and Øverland, G. 2016. *Responding to Global Poverty.* Cambridge: Cambridge University Press.

Beerbohm, Eric. 2012. *In Our Name: The Ethics of Democracy.* Princeton, NJ: Princeton University Press.

Bernstein, S. 2017. “Causal Proportions and Moral Responsibility.” In *Oxford Studies in Agency and Responsibility*, vol. 4, edited by D. Shoemaker, 165–182. Oxford: Oxford University Press.

Bernstein, S. Forthcoming. “David Lewis’ Theories of Causation and Their Influence.” In *Cambridge History of Philosophy*, edited by K. Becker.

Braham, M., and van Hees, M. 2009. “Degrees of Causation.” *Erkenntnis* 71, no. 3: 323–344.

Brennan, G., and Lomasky, L. 1993. *Democracy and Decision: The Pure Theory of Electoral Preference.* New York: Cambridge University Press.

Brennan, Jason. 2016. “The Ethics and Rationality of Voting.” In *Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Accessible at: **https://plato.stanford.edu/entries/voting/**

Broome, J. 2012. *Climate Matters.* New York: W. W. Norton.

Budulfoson, M. B. Forthcoming. “The Inefficacy Objection to Consequentialism and the Problem with the Expected Consequences Response.” *Philosophical Studies*.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Bunzl, M. 1979. “Causal Overdetermination.” *Journal of Philosophy* 76, no. 3: 134–150.

Cullity, G. 2000. “Pooled Beneficence.” In *Imperceptible Harms and Benefits*, edited by M. Almeida, 9–42. Dordrecht, the Netherlands: Kluwer.

Dowe, P. 2000. *Physical Causation.* Cambridge: Cambridge University Press.

Edlin, A., Gelman, A., and Kaplan, N. 2007. “Voting as a Rational Choice: Why and How People Vote to Improve the Well-Being of Others.” *Rationality and Society* 19:219–314.

Eriksson, Anton. Unpublished manuscript. “Omitting to Emit.” PhD diss., University of Sheffield.

Fenton-Glynn, L. 2013. “Causal Foundationalism, Physical Causation, and Difference-Making.” *Synthese* 190, no. 6: 1017–1037.

Garvey, J. 2011. “Climate Change and Causal Inefficacy: Why Go Green When It Makes No Difference?” *Royal Institute of Philosophy Supplement* 69: 157–174.

Gesang, B. 2017. “Climate Change—Do I Make a Difference?.” *Environmental Ethics* 39, no. 1: 3–19.

Glover, J., and Scott-Taggart, M. 1975. “It Makes No Difference Whether or Not I Do It.” *Proceedings of the Aristotelian Society, Supplementary Volumes* 49:171–209.

Gunnemyr, M. 2019. “Causing Global Warming.” *Ethical Theory and Moral Practice,* 22, no.2: 399–424.

Hiller, A. 2011. “Climate Change and Individual Responsibility.” *The Monist* 94, no. 3: 349–368.

Hitchcock, C. R. 1993. “A Generalised Probabilistic Theory of Causal Relevance.” *Synthese* 97, no. 3: 335–364.

Jackson, F. 1987. “Group Morality.” In *Metaphysics and Morality: Essays in Honour of J.J.C Smart*, edited by P. Pettit, R. Sylvan and J. Norman, 91–110. Oxford: Blackwell.

Kagan, S. 2011. “Do I Make a Difference?” *Philosophy and Public Affairs* 39, no. 2: 105–141.

Kahn, B. Forthcoming. *Global Poverty, Structural Injustice, and Collectivization.*

Kaiserman, A. 2016. “Causal Contribution.” *Proceedings of the Aristotelian Society* 116, no. 3: 387–394.

Kingston, E., and Sinnott-Armstrong, W. S. 2018. “What’s Wrong with Joyguzzling?” *Ethical Theory and Moral Practice* 21, no. 1: 169–186.

(p. 652) Kment, B. 2010. “Causation: Determination and Difference Making.” *Nous* 44, no. 1: 80–111.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Lawford-Smith, H. 2016a. “Climate Matters *Pro Tanto*, Does It Matter All-Things-Considered?” *Midwest Studies in Philosophy* XL: Ethics and Global Climate Change 40, no. 1: 129–142.

Lawford-Smith, H. 2016b. “Difference-Making and Individuals’ Climate-Related Obligations.” In *Climate Justice in a Non-Ideal World*, edited by C. Hayward and D. Roser, 64–82. Oxford: Oxford University Press.

Lawford-Smith, H. 2018. “Does Purchasing Make Consumers Complicit in Global Labour Injustice?” *Res Publica* 24, no. 3, 319–338.

Lawford-Smith, H. 2019. *Not In Their Name: Are Citizens Culpable for Their States’ Actions?* Oxford: Oxford University Press.

Lewis, C. 2000. “Causation as Influence.” *Journal of Philosophy* 97, no. 4: 182–197.

Lewis, D. 1973. “Causation.” *Journal of Philosophy* 70, no. 17: 556–567.

Lewis, D. 1986. “Causation.” Reprinted in *Philosophical Papers, Vol. II*, 159–213. New York: Oxford University Press.

Lewis, D. 2004. “Causation as Influence.” In *Causation and Counterfactuals*, edited by J. Collins, N. Hall and L. Paul, 75–106. Cambridge, MA: MIT Press.

MacAskill, Will. 2015. *Doing Good Better.* New York: Penguin Books.

McGrath, S. 2003. “Causation and the Making/Allowing Distinction.” *Philosophical Studies* 114, no. 1–2: 81–106.

Menzies, P. 2004. “Difference-Making in Context.” In *Causation and Counterfactuals*, edited by J. Collins, N. Hall, and L. Paul, 139–181. Cambridge, MA: MIT Press.

Morgan-Knapp, C., and Goodman, C. 2015. “Consequentialism, Climate Harm, and Individual Obligations.” *Ethical Theory and Moral Practice* 18, *no.* 1: 177–190.

Nefsky, J. 2012. “Consequentialism and the Problem of Collective Harm: A Reply to Kagan.” *Philosophy and Public Affairs* 39, no. 4: 364–395.

Nefsky, J. 2017. “How You Can Help, without Making a Difference.” *Philosophical Studies* 174, no. 11: 2743–2767.

Nefsky, J. 2019. “Collective Harm and the Inefficacy Problem.” *Philosophy Compass* 14, no. 4: e12587.

Nolt, J. 2011. “How Harmful Are the Average American’s Greenhouse Gas Emissions?” *Ethics, Policy and Environment* 14, no. 1: 3–10.

Norcross, A. 2004. “Puppies, Pigs, and People: Eating Meat and Marginal Cases.” *Philosophical Perspectives* 18, no. 1: 229–245.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Nye, A. 2009. “Physical Causation and Difference-Making.” *The British Journal for the Philosophy of Science* 60*, no.* 4: 737–764.

Parfit, D. (1984) 2003. “Five Mistakes in Moral Mathematics.” In *Reasons and Persons*. 68–86. Oxford: Oxford University Press.

Pellegrino, G. Forthcoming. “Robust Individual Responsibility for Climate Harms.” *Ethical Theory and Moral Practice*.

Pinkert, F. 2015. “What If I Cannot Make a Difference (and Know It).” *Ethics* 125, no. 4: 971–998.

Quinn, W. 1990. “The Puzzle of the Self-Torturer.” *Philosophical Studies* 59: 79–90.

Rendall, M. 2015. “Carbon Leakage and the Argument from No Difference.” *Environmental Values* 24, no. 4: 535–552.

Sandberg, J. 2011. “My Emissions Make No Difference.” *Environmental Ethics* 33, no. 3: 229–248.

Sartorio, C. 2005. “Causes as Difference-Makers.” *Philosophical Studies* 123: 71–96.

(p. 653) Sartorio, C. 2007. “Causation and Responsibility.” *Philosophy Compass* 2, no. 5: 749–765.

Sartorio, C. 2010. “Causation and Ethics.” In *The Oxford Handbook of Causation*, edited by H. Beebee, C. Hitchcock, and P. Menzies, 576–589. Oxford: Oxford University Press.

Schaffer, J. 2003. “Overdetermining Causes.” *Philosophical Studies* 114, no. 1–2: 23–45.

Schwenkenbecher, A. 2014. “Is There an Obligation to Reduce One’s Individual Carbon Footprint?.” *Critical Review of International Social and Political Philosophy* 17, no. 2: 168–188.

Singer, Peter. 2015. *The Most Good You Can Do.* New Haven, CT: Yale University Press.

Sinnott-Armstrong, W. 2005. “It’s Not My Fault: Global Warming and Individual Moral Obligations.” In *Perspectives on Climate Change*, edited by W. Sinnott-Armstrong and R. Howarth, 221–253. Amsterdam: Elsevier.

Smart, J. J. C., and Williams, B. 1973. *Utilitarianism—For and Against.* Cambridge: Cambridge University Press.

Spiekermann, K. 2014. “Small Impacts and Imperceptible Effects: Causing Harm with Others.” *Midwest Studies in Philosophy* 28: 75–90.

Talbot, B. 2018. “Collective Action Problems and Conflicting Obligations.” *Philosophical Studies* 175, no. 9: 2239–2261.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Vance, C. 2017. “Climate Change, Individual Emissions, and Foreseeing Harm.” *Journal of Moral Philosophy* 14, no. 5: 562–584.

Vanderheiden, A. 2007. “Climate Change and the Challenge of Moral Responsibility.” *Journal of Philosophical Research* 32 (Suppl.): 85–92.

Waters, C. K. 2007. “Causes That Make a Difference.” *The Journal of Philosophy* 104, no. 11: 551–579.

Weslake, B. Forthcoming. “Difference-Making, Closure and Exclusion.” In *Making a Difference*, edited by H. Beebee, C. Hitchcock and H. Price. Oxford: Oxford University Press.

Williamson, J. 2010. “Probabilistic Theories.” In *The Oxford Handbook of Causation*, edited by H. Beebee, C. Hitchcock and P. Menzies, 186–210. Oxford: Oxford University Press.

Wright, R. 1985. “Causation in Tort Law.” *California Law Review* 73, no. 6: 1788–1813.
(p. 654)

Holly Lawford-Smith

Holly Lawford-Smith is a Senior Lecturer in Political Philosophy at the University of Melbourne. She works on topics across moral and political philosophy, applied ethics, and social ontology, including climate ethics, corporate responsibility, collective agency, and radical feminism. Her first book Not in Their Name: Are Citizens Culpable for Their States’ Actions? came out with Oxford University Press in 2019.

William Tuckwell

William Tuckwell is a PhD candidate in Philosophy at The University of Melbourne. His research focuses mainly on social and political philosophy, epistemology, and the interconnections between the two.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The Love–Hate Relationship between Feminism and Consequentialism

Samantha Brennan

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.30

Abstract and Keywords

The early utilitarians were strong champions for the equal treatment of women, yet contemporary feminists are some of consequentialism's biggest critics. Arguing from a more generous account of what counts as consequentialist moral reasoning, this chapter identifies feminist criticisms of consequentialism and sees whether, and to what extent, feminism and consequentialism can be reconciled. It argues that a feminist version of consequentialism is possible and, regardless, that all feminist moral theories contain significant consequentialist elements which it would be a mistake to ignore. Finally, it suggests that all feminist approaches to ethics ought to accord some role to consequences and results, and therefore ought to contribute to debates and discussions within consequentialist ethics.

Keywords: feminism, consequentialism, relational ethics, utilitarianism, well-being, impartialism, desire theory, preferentism, perfectionism, Kantian deontology

1. Introduction

FEMINISM, insofar as it is concerned with bringing a better world for all, one in which women are afforded equal opportunities to flourish and everyone is freed from gender-based oppression, has always had consequentialist elements that form an important part of the theory.[1] When feminists compare one country to another country, or our world to some future or past world, we look to people's lives and evaluate them in terms of equality, well-being, and freedom. Feminist utopian fiction, feminist political activism, and feminist moral philosophy all share the idea that one important way in which a world can be a better world is if women are treated with equal respect and have equal opportunities. Feminist moral and political philosophy draws a picture of a better, more equal world and evaluates actions and policies in terms of their contribution to bringing that world about.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(p. 617) Consequentialism, insofar as it is concerned with the well-being of all persons, or total well-being however measured, has historically paid particular attention to women, because of the well-being gap between women and men. Improving the life prospects of girls and women remains a focus of many consequentialist approaches to applied moral philosophy (Nussbaum 1995; 2000). Improving the educational prospects of girls and women turns out to be a remarkably easy way to raise overall levels of well-being in developing countries (Sen 1990; 2000; Duflo 2012). Much feminist moral reasoning is consequentialist, and much consequentialist reasoning is focused on the well-being of girls and women. The two approaches are intertwined in fundamental ways.

It's my view that to say we have a moral reason to do x because x would make the world better is to engage in consequentialist moral reasoning. One might want a more restrictive definition of what makes a theory consequentialist all things considered—perhaps this consideration must be the most significant. But to ignore the consequentialist elements in most forms of moral reasoning leaves us in danger of making serious mistakes in moral theorizing. Insofar as all reasonable moral theories require us to promote the good some or most of the time, we all need to work on the shape and strength of those requirements and the nature of the good that is to be promoted.

We can see this deep connection when we look at some similarities that emerge between feminist and consequentialist moral reasoning. Here are two of them.

First, both theories regard all actions, indeed all of our lives, as up for moral/political scrutiny and evaluation. "The personal is political" is a slogan associated with second-wave feminist movements in North America and Europe. All aspects of our lives are up for political examination, even things we might think of as private hobbies or habits. Consequentialism, too, looks at all aspects of our lives, and all of our actions are up for moral scrutiny. Indeed, this scope question is sometimes thought to be one of the ways in which consequentialism is too morally demanding. Feminists have raised this objection to consequentialism (Driver 2005), but both feminism and consequentialism present ways of seeing the world that can be all-encompassing, demanding, and at odds with our everyday or default ways of viewing the world.

Second, in a related vein, both feminism and consequentialism address practical moral problems and take on questions in everyday ethics. In both these communities, one hears lengthy, heated debates about parenting ethics, sexual ethics, the ethics of consumption, and the morality of eating meat, to give just a few examples. It is no surprise given the scope question earlier. The earliest philosophical defense of same-sex (male/male) sexual relationships,[2] animal rights,[3] and the moral status of prisons[4] came from the early utilitarians. Today we find some, but not all, feminists and consequentialists making very similar arguments against legal restrictions on sex work[5] and abor (p. 618) tion.[6] Both sets of arguments look to the results of criminalization and reason from principles of harm reduction to legalization. The shared focus is on the effects of policies on the lives of persons.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Despite considerable common ground, feminism and consequentialism have enjoyed a very long love–hate relationship. While some of the earliest defenders of feminism were utilitarians, and important consequentialist elements sit at the core of contemporary feminism, today some of the harshest critics of consequentialist moral and political reasoning are feminists. In this essay, I set out to tell the story of the connection between consequentialism and feminism and in doing so hope to sort out some of the areas of disagreement. I also hope to show that there is room for a feminist version of consequentialism and to explain why, regardless of whether they are in complete agreement, all feminists ought to pay attention to consequentialist moral theory.

Annette Baier criticizes mainstream moral theory as incomplete. She writes, “Most of what are recognised as the current moral theories are also incomplete, since they do not purport to be yet really comprehensive. Wrongs to animals and wrongful destruction of our physical environment are put to one side … and in most ‘liberal’ theories there are only hand waves concerning our proper attitude to our children, to the ill, to our relatives, friends and lovers” (Baier 1985, 55). This surely isn’t true of consequentialist ethics.

2. The Historical Connection

We can consider the historical connection between consequentialism and feminism by looking at the writing of John Stuart Mill. I want to look at Mill’s feminist utilitarianism because one of the problems that confronts Mill’s theory is still an issue for feminists and consequentialists today. What status should desires developed under patriarchy have? How you think about Mill’s answer is connected to what kind of feminist you take Mill to be, where liberal feminists take women’s individual desires at face value, and radical feminists look at the social contexts in which those desires were developed to critically engage with them.[7] There are, of course, other differences between liberal and radical feminism, but to the extent to which Mill looks to the coercion that shapes women’s plans, preferences, and desires, it appears that, on this issue at least, he is on the side of fairly radical versions of feminism.

The background to Mill’s work will be familiar to most of us. When Mill wrote *The Subjection of Women* in 1869, women had no access to education, and there was a strict system of gender role socialization. Women had very few rights in marriage, only limited rights to work, no political or voting rights, and very limited property rights and no right to inherit. In the face of this widespread inequality between men and women, Mill took the comparative status of men and women to be one of the most important (p. 619) issues for moral and political philosophy. Mill argued for women’s rights, women’s suffrage, and equal access to education for women. In *The Subjection of Women*, claimed by Mill to be coauthored with his spouse, he defended equality between the sexes.[8] Mill thought the unequal treatment of women under the law both wronged individual women and set back progress on a social level. Treating women as lesser than men, at least to the extent that they were treated unequally in his day, was on Mill’s view inconsistent with utilitarian

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

thinking. Individually and collectively we miss out on the contributions to society that women might have to make. Both men and women have their personal moral character harmed by the injustice of unequal social arrangements (Okin and Mansbridge 2005). Writes Mill (1984, 261), “That the principle which regulates the existing social relations between the two sexes—the legal subordination of one sex to the other—is wrong itself, and now one of the chief hindrances to human improvement; and that it ought to be replaced by a principle of perfect equality, admitting no power or privilege on the one side, nor disability on the other.”

With barriers to women’s participation in the workplace and political life removed, Mill still imagined that women would choose to stay at home and care for children. Utilitarian moral reasoning required that other options be available to women even if very few women selected them. We are happier and more fulfilled when the work we take on, whether inside or outside the home, is taken on by choice. But why think women, rather than men, would choose family over work and politics? Mill was far ahead of others writing about women’s education in that he saw a connection between the education of girls and women and judgments about what roles and activities suited the female sex. He believed that we cannot claim to know the true nature of women because women are what men have made them to be. Writes Mill (273): “What is now called the nature of women is an eminently artificial thing—the result of forced repression in some directions, unnatural stimulation in others.” People tend to think that whatever is customary is natural, especially those who have the power under the customary arrangement, according to Mill (267): “But was there ever any domination which did not appear natural to those who possessed it?”

Mill recognized the forces at work that encouraged women to develop strong desires to be found attractive by men. Calling the object of being found attractive by men the “polar star of feminine education and formation of character” (272), Mill attributes it to the rewards at work. Women have limited career options if they do not marry, and if they do marry, they are financially dependent on their husbands. Mill recognized that what many thought of as the nature of women was artificial, the result of an upbringing focused on marriage.

The subordination of women involved education because of the kind of subordination it is. According to Mill, women’s oppression in marriage and the family begins with the ways in which the characters of girls are formed.

(p. 620)

> Men do not want solely the obedience of women; they want their sentiments. All men, except the most brutish, desire to have, in the woman most nearly connected with them, not a forced slave but a willing one, not a slave merely, but a favourite. They have therefore put everything in practice to enslave their minds. The masters of all other slaves rely, for maintaining obedience, on fear; either fear of themselves, or religious fears. The masters of women wanted more than simple obedience, and they turned the whole force of education to effect their purpose. All

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> women are brought up from the very earliest years in the belief that their ideal of character is the very opposite to that of men; not self will, and government by self-control, but submission, and yielding to the control of others. (Mill 1984, 271)

Having outlined these features of Mill's feminism, we can now ask what sort of feminist Mill is. The standard view is that Mill is a liberal feminist. The reforms he advocates are legal ones. The barriers to women's education were, in his day, a matter of rule and policy. The main point of Mill's reform was to change the role of women in marriage, to make men and women partners in married life.

There is also an optimism in Mill's writing, a belief in moral progress and in rational change. He envisioned marriage as a partnership between equals. Once women were equal partners in marriage and received a reasonable education, other barriers to women's political equality would fall away, thought Mill.

But aspects of Mill's feminism go beyond standard liberal feminism. His emphasis on patriarchal society as a holdover from less rational times sounds much more like radical feminism. His emphasis on violence, power, and the physical force involved in women's subordination also sounds more like radical than liberal feminism. Mill also connects oppression in the home with women's exclusion from politics and public life. He argues that it is the role women play in marriage that is fundamental to understanding gender oppression. Writes Keith Burgess-Jackson (1995, 369), "What puzzles is that Mill's views on the social and legal status of women are more closely aligned with those of contemporary radical feminists than with those of contemporary liberal feminists."

Is there a conflict between Mill's feminism and his liberalism? Will solutions available to Mill address the problem as he understands it? This problem cuts two ways for Mill. First, on his view, law as a tool of reform is limited by the harm principle (so sometimes we cannot use the law when doing so might help). Second, extralegal measures don't match up with Mill's focus on legislative reform. Of course, in Mill's times the main barriers to women's participation *were* legal. The need for such barriers formed part of his argument that sex differences were not natural. If they were natural, laws would not be needed. Mill might not have imagined how many of the patriarchal norms of marriage and family life could persist without legal enforcement.

There is also thought to be a problem accounting for women's agency in oppression. Why is it that more women don't rebel? Mill's answer is that unlike standard slavery by force, patriarchal oppression begins by taking over women's minds, making them willing slaves. Suppose that women, many of them at any rate, are willing slaves—as adults (p. 621) they choose or would endorse the conditions of their oppression. How then do we respect their choices? Should women be treated as having moral agency given an upbringing which stunts our capacities?

The worry is that a feminist understanding of the creation of selves in a patriarchal society may undermine treating women as autonomous. For on Mill's understanding of the role of education and upbringing in shaping our characters, it's not just our desires that

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

might be mistaken. Indeed, it's the self underneath those desires which may be misshapen. Compare the world of pornography in which women are paid (in a world in which women's looks are more highly valued than women's intellects, participating in porn as work can be rational) versus the world of *volunteer* pornography. There is a sense in which the new world is better. No one is forced, not even through financial coercion. Arguably there is an increased emphasis on women's sexuality and on female pleasure. But it's also parallel to Mill's notion of wives as willing slaves. Better from the point of view of men that women *choose* to take off their clothes on the beach for the *Girls Gone Wild* cameras.

Here are two possibilities: Patriarchal society creates the characters of women. Female socialization and gendered education create women as willing participants in traditional marriage and in porn culture. On this way of thinking about women's desires, Mill is a radical feminist. It's consistent with Mill's emphasis on violence and coercion and on his characterization of the family as an incubator for social injustice. Or we might think that women autonomously choose to play the role of traditional wife and willing porn model, and given that these choices are not made in coercive circumstances, we must respect them. It is true that women's desires are responsive to the rewards and benefits in the society that surrounds them, but this is true for all desires.

Alternatively we could give up versions of utilitarianism that build on subjective welfare and go the perfectionist route instead. On this view, we have duties to ourselves as women not to choose lives which do not allow us to flourish as whole persons. Feminist game theorists might look at the problem in decision-theoretic terms. It might be in the interests of an individual woman to pursue certain desires, but it could also be collectively self-defeating for women.

Why recap Mill's version of feminist utilitarianism? In what follows I want to present and assess a range of feminist worries about consequentialist moral theory. It will turn out that one of the most serious was the question about the status of selves and the desires that are formed in a patriarchal society.

3. Feminist Criticisms of Consequentialist Moral Reasoning

Feminist moral theory begins with criticisms of mainstream moral philosophy, both for its method and for what it excludes: women's experiences and intuitions and often areas of life of concern to women. Mainstream ethics is said to be individualistic, abstract, (p. 622) idealizing (based on false assumptions about persons, including that we are all rational, independent, and autonomous), and universalizing. Feminist work in ethics either proceeds by developing new approaches to ethics, revising existing theories, or abandoning theory-driven approaches altogether and working bottom-up from lived ethical experience and real-world moral problems.

Feminist criticisms of consequentialist moral reasoning are particularly interesting because they are rarely about the conclusions reached. There is a small industry of raising objections to consequentialism on the basis of the results of thought-experimental coun-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

terexamples, such as the trolley problem.[9] Given that feminist ethical thinking is grounded in nonideal theory and actual lives in real-world circumstances, consequentialism stands a better chance. After all, most consequentialist arguments in ethics support politically progressive conclusions. Indeed, some consequentialists think the political realm is the area of life to which consequentialism is best suited (Goodin 1995). Yet few feminists endorse explicitly consequentialist moral theories. Why is that? Consequentialist arguments might reach the right conclusions, but they do it in a way that is not particularly feminist. It is not enough for a moral theory to arrive at the intuitively right answer. It must also do it in the right way (Brennan 1999, 860).

Here I will outline four feminist worries about consequentialist moral reasoning which focus on method rather than outcome.

Worry 1: Is happiness or well-being morally weighty enough to ground women's equality? Consequentialists might argue against the continued subordination of women, but they usually do it on the basis of the total amount of happiness or well-being, not on the basis of the wrongs that constitute that subordination. Even if the version of consequentialism at issue is more sophisticated than utilitarianism, it still might leave out consideration of moral wrongs. The arguments based on consequences alone are not explicitly feminist. Consider Mill's arguments for equality in marriage, in the workplace, and in education. His claim was that we are all worse off as a result. This seems too feeble a basis for such an important moral claim. Likewise, consider a case in which the unequal treatment in fact proposed more good overall. Some philosophers think this might be true when we consider maintaining sexist gender norms, such as door opening. The norms have considerable utility and play an important role in social coordination. Changing norms, even sexist norms, comes with a cost. If that were true, then utilitarianism would yield the result that continued subordination in some cases was morally preferable. That women deserve equal treatment seems to be something we can know without making calculations of overall utility or maximum happiness. Equality comes before utility in feminist moral theory.

How might consequentialists respond to this concern? Consequentialist moral theories can include egalitarian considerations as central. There are three different places in consequentialist moral theory where equality can play a role. First, for all consequentialist moral theories, it is true that people's well-being matters equally. The equality of persons is foundational to consequentialist moral reasoning. This does not guarantee equality of outcome. Second, equality can matter instrumentally because quite often the (p. 623) distribution of goods that produces the most utility is the most equal distribution.[10] Third, a version of consequentialism with a pluralist conception of the good might well include equality as part of the good to be maximized. While it is true that such a pluralist account of the good has hard questions to answer—What is equality and how do we measure it? How ought we to trade off equality with the other goods that matter? (Temkin 1986)—it is a conceptually possible and attractive version of consequentialism that might be more amenable to feminist theorizing.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

One might worry that these replies still don't seem robust enough. Suppose men are utility monsters who get far more utility oppressing women than women get disutility from being oppressed and so much so that their utility outweighs even the noninstrumental disvalue in the resulting gender inequality. In general there are two kinds of responses one can make to this worry. First, one can make the claim, as many consequentialists do, that our self-regarding preferences are always stronger than the preferences we have regarding the treatment of others. For example, I might be a homophobic person who experiences disutility at the sight of seeing same-sex couples kiss. Now imagine a policy that restricted the rights of same-sex couples to engage in public displays of affection. It seems likely the disutility of not being able to live one's life as one chooses is far stronger than my preferences regarding how others behave. Second, one can limit the amount of utility that is considered for each person, thus imposing a kind of structural equality on the good that is to be maximized.

Worry 2: Feminist approaches to ethics tend to focus not on outcomes but rather on the method by which we arrive at a particular outcome. We can see this in care ethics and in other relational approaches to ethical theorizing. According to this line of criticism, consequentialism goes wrong when it evaluates acts and policies solely on the basis of outcomes. Instead, in order to evaluate an act or a policy ethically, we need to know how a particular outcome came about. Often one hears this criticism of consequentialist moral reasoning from rights theorists or from libertarian political philosophers who echo Robert Nozick's (1974, 160–164) emphasis on historical versus patterned principles of justice. But the point is generalizable beyond rights-based theories.

Feminist moral theorists who give special significance to care and to relationships might also find it narrow just to look at results. For example, Virginia Held writes, "Utilitarians suppose that one highly abstract principle, the Principle of Utility, can be applied to every moral problem no matter what the context. A genuinely universal or gender-neutral moral theory would be one that would take account of the experience and concerns of women as fully as it would take account of the experience and concerns of men. When we focus on women's experience of moral problems, however, we find that they are especially concerned with actual relationships between embodied persons and with what these relationships seem to require" (Held 1990, 330).

What would an account of ethics that made relationships central look like? For one example of relational thinking, we can look to Susan Sherwin's feminist analysis of abortion. Because the life of the fetus is dependent on the pregnant woman, Sherwin (1991, (p. 624) 109) thinks that "fetuses are morally significant but their status is relational rather than absolute." She writes, "Because humans are fundamentally relational beings, it is important to remember that fetuses are characteristically limited in the 'relationships' in which they can 'participate'; within those relationships, they can make only the most restricted 'contributions'" (Sherwin 1991, 110).

Nel Noddings thinks that maintaining caring relationships is a universal value:

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> A and B, struggling with a moral decision, are two different persons with different factual histories, different projects and aspirations and different ideals. It may indeed be right, morally right, for A to do X and B to do not-X. We may, that is, connect right and wrong to the ethical ideal. This does not cast us into relativism, because the ideal contains at its heart a component that is universal: Maintenance of the caring relation. (Noddings 1984, 85–86)

Recent work in feminist ethics has seen a broadening of the scope of concern to include issues such as trust (McLeod 2002), moral dilemmas (Tessman 2017), autonomy (Mackenzie and Stoljar 2000), and responsibility (Card 1996). As well as broadening the scope this way, feminist ethics has also applied relational thinking to other concepts such as personhood and rights. In addition to applying relational insights to moral concepts, feminists have argued that persons themselves are best understood in relational terms. And so moral concepts such as autonomy, rights, and freedom have undergone a relational transformation. Feminists have also expanded the range of relationships to which ethical demands have been thought to apply. Feminists have been critical of most traditional approaches to ethics as focusing too much on the adult, independent, autonomous person in his interactions with other persons sharing those same characteristics.

But a feminist who wanted to revise consequentialism to account for the relational insights of feminist ethics might think that we could include valuable relationships as part of the good to be promoted, either because relationships are instrumentally valuable or because they matter in and of themselves. Is recognizing the intrinsic value of relationships enough to respond to this worry about consequentialist moral reasoning? Does it value relationships in the right way? For example, on an agent-neutral version of consequentialism, one will be required to betray one relationship for the sake of minimizing one's relationship betrayals. But some wonder whether a true relationship can be valued only agent-neutrally such that one would be willing to sacrifice that relationship for the sake of minimizing the number of relationships that one sacrifices. Perhaps, being willing to sacrifice a relationship for the greater good is incompatible with having the most valuable sort of relationship. Should I be willing to sacrifice my relationship with my daughter even if this would be the only way to prevent several others from sacrificing their relationships with their daughters?[11]

So it may be that an account of consequences which is all about agent-neutral values is insufficient to respond to the relational insights of feminist ethics. But even a more (p. 625) sophisticated version of consequentialism which incorporates agent-relative values is still results focused. There may be versions of feminist ethics according to which it matters *how* an outcome comes about, and for such approaches to feminist ethics, a focus on consequences cannot be the whole story.

As I will claim later though, consequences are still a very important part of the story, and to that extent all feminist approaches to ethics are consequentialist.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Worry 3: Perhaps the strongest and most frequent charge against consequentialist moral reasoning made by feminists is against impartialism, the aspect of consequentialist reasoning that says each person's good counts for the same. Should I read bedtime stories to my children or read to the children who would benefit the most from being read to? Do I visit my elderly relatives in the hospital or visit those for whom my visiting would promote the most utility? Surely relations matter more and our ethical life ought to give pride of place to close connections to family and friends? Impartiality isn't just about how we regard outcomes. Some utilitarians think it exemplifies the moral perspective.[12] Peter Singer, for instance, says, "My ability to reason shows me the possibility of detaching myself from my own perspective and shows me what the universe might look like if I had no personal perspective" (Singer 1993, 229).

Asks Julia Driver in the course of her response to the feminist partialist criticisms of consequentialism, "What kind of ethics would really prescribe a disinterested attitude here? We ought to pay more attention to our children than to humanity as a whole. We owe our family and friends more consideration, and ideals of family and friendship hold that it is perfectly okay, morally, to do what one wants to do, to go along with one's feelings—and that, indeed, this can even be morally better than taking the god's-eye view of one's actions" (Driver 2005, 3).

Helga Kuhse, Peter Singer, and Maurice Rickard put the contrast between the two approaches this way: "Partial moral reasoning is central to the care orientation, involves judgments that emphasize personal relationships and attachments. These sorts of judgments and dispositions differ from impartialist judgments in that they favour people with whom we are personally connected over people with whom we are not. Impartialist reasoning, by contrast, is central to standard moral thinking, and involves judgments and dispositions that are detached and do not favour personal attachments. They reflect concern for what equal consideration of people's interests requires, as well as wider impersonal responsibilities" (Kuhse, Singer, and Rickard 1998, 453).

In "A Feminist Approach to Ethics," Susan Sherwin writes that impartiality is a requirement of utilitarian approaches to ethics that puts our intuitions about who to care the most for at odds with duty-based reasoning. She writes, "But any particular decision will still depend on accidental empirical facts, and it may often turn out that benefitting strangers or even enemies creates more utility than any alternative, despite the disutility of our possible distaste for such a situation. If we accept utilitarianism, it is our duty to concentrate on how we can produce the greatest utility with only incidental and (p. 626) instrumental concern for who is being benefitted or harmed. In the final analysis, we must be impartial in increasing utility despite our prior preferences" (Sherwin 1984, 706).

Here I present three responses to the worry about impartialist reasoning being at odds with our everyday, lived partialist moral experience.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

First, we can note that while feminists raise these criticisms, there is nothing necessarily feminist about them. Many other moral theorists, not writing from a feminist perspective at all, have criticized utilitarianism for its inability to account for close, personal relationships and its overly demanding nature both in terms of scope and strength of obligation.[13]

Second, it's also not just a criticism of consequentialism. Kantian deontology has its own issues with impartialism, which arguably are worse since they require acting from the motive of impartialist duty. For a Kantian, it is not enough that the act bring about the result which is justified on the basis of impartialist reasons.

Third, sophisticated, two-level versions of consequentialist theory need not be committed to impartialism at the level of individual moral decisions. One can draw a distinction between consequentialism as the ground of right action and consequentialism as a decision procedure (Driver 2005, 192).

Kuhse and her coauthors, for example, argue that impartialist and partialist reasoning are compatible but separate levels of moral thinking, impartialist reasoning being at the critical level and partialist reasoning being at the intuitive. They write,

> According to this view, correct moral outcomes or decisions would be arrived at through principled, abstract, impartialist reasoning. If an outcome or decision cannot be justified in terms of equality, reciprocity, merit, or other impartialist considerations, then it is not the correct outcome or decision. On some sorts of occasion, the impartialist critical level thinking is possible and appropriate. For instance, in public life, it is usually possible to take into account all the relevant factors, and to justly and equitably arrive at appropriate decisions and outcomes. On many other occasions, however, the time, energy and information are just not available for extended impartialist reasoning. (Kuhse et al., 1998, 460)

It may be the case that consequentialism is the right answer to what makes acts right or wrong, but that one doesn't have to have consequentialist considerations front of mind when making moral choices. Indeed, in some areas of life, it might be better from a consequentialist perspective if we did not act as consequentialists. This shows that a feminist version of consequentialism would need to be a sophisticated, two-level version of consequentialism, but many consequentialists think this is the most plausible version of their moral theory anyway.

Worry 4: A particularly feminist criticism of consequentialism focuses on consequentialism's theory of the good. What makes a person's life go well? One common answer is that well-being consists in the satisfaction of one's desires. But feminists have (p. 627) raised serious worries about desires that develop in a patriarchal society, calling them "false" or "deformed" desires. The worry about such desires is that it does not seem plausible that their satisfaction improves things for the persons who have them or for the state of the world. This worry grapples with the concern we raised for Mill's account of women's desires.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Sandra Bartky writes about false desires that “fasten us to the established order of domination, for the same system which produces false needs also controls the conditions under which such needs can be satisfied. ‘False needs,’ it might be ventured, are needs which are produced through indoctrination, psychological manipulation, and the denial of autonomy; they are needs whose possession and satisfaction benefit not the subject who has them but a social order whose interest lies in domination” (Bartky 1990, 42; cited in Superson 2005).

Of course, it is natural to shape our desires to fit the world around us, but when our social context is built of norms that are sexist, racist, ableist, and so forth, our desires then reflect back that social reality. Writes Superson, “Arguably all desires are formed in a social context; deformed desires are formed by unjust social conditions, including those where men are deemed superior and women inferior” (Superson 2005, 109).

Let me present a simple example that I often use when teaching but have not until now shared in my writing. As an undergraduate student, I loved my philosophy classes. I was drawn into the world to which my philosophy professors introduced me. I enjoyed arguments and analysis, reading difficult texts, and writing philosophy papers. I fell in love with philosophy, as many of us do. I also imagined a future in which I was a philosopher’s girlfriend. That was the life I wanted! All of my future fantasies about being a philosopher’s girlfriend involved staying up all night arguing about ideas. This should have been a clue that I didn’t really have a crush on my philosophy professors. It was only when one of those same professors suggested that I apply to graduate school that my “philosopher’s girlfriend” desires went away. I didn’t need to date a philosopher. I could become one. My world had opened up, and my desires changed to account for the new possibilities.

Martha Nussbaum has focused on the desires that women develop in very traditional, patriarchal societies. She thinks there are three general factors present in patriarchy that produce deformed desires: (1) lack of information or false information about facts, (2) lack of reflection or deliberation about norms, and (3) lack of options (Nussbaum 1999, 149). Likewise, Serene Khader writes, “People’s wants can become deformed by bad circumstances. Taking preference satisfaction as the end of development implies that we have an obligation to fulfill people’s deformed preferences. This implication of preference satisfaction theories of social distribution is deeply objectionable … and so we should be wary of utilitarian approaches” (Khader 2011, 43).

Superson argues that such desires are irrational, an affront to an agent’s autonomy, and ought to be rejected. At the very least we should not build a moral theory out of their satisfaction. But not all feminists reject desire-based accounts of the good. Some feminists think we can revise the desire theory such that the desires whose satisfaction improves a person’s well-being (and the overall good) aren’t the person’s actual desires. (p. 628) Instead, the desires whose satisfaction matters morally are those that would survive various “improvements” such as the addition of more information or greater rationality.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Other feminists such as Harriet Baber argue for the preference account—as distinct from desire theory—on feminist grounds. Writes Baber,

> Preference utilitarianism, I have argued, is good for women. Preferentism provides an account of utility as relative well-being according to which impoverished women in the Global South are very much worse off than affluent individuals in the Global North—not because they suffer from "deformed" or "inappropriately adaptive" preferences but because they are harmed by deformed social, economic, and political arrangements that prevent them from attaining goods that are higher on their preference rankings. That, intuitively, is unfair. Preferentism, while it does not provide an account of fairness, is a starting point for formulating such an account, and for the discussion of how unjust institutions and unfair social arrangements can be dismantled in order to provide people with the widest possible range of options for preference satisfaction so that utility may abound. (Baber 2017, 19)

What's the distinction between the desire theory and a preference-satisfaction account? According to Baber, "The 'desire theory' is an account of well-being according to which only those states of affairs that a person desires contribute to her well-being. Preferentism, however, is an account of relative well-being. Preferentism ranks states of affairs according to their relative betterness" (Baber 2017, 3).

So we have a range of possible feminist options. We can revise the desire theory to take account of feminist objections, or we can move to a preference account as Baber suggests. Still some feminists argue that we need to reject all subjective accounts of the good. Do we need to move to an objective theory of the good? Kimberly Yuracko (2003) argues for a feminist version of perfectionism as the best basis for feminist arguments for women's autonomy. Martha Nussbaum's capacity-based account of human rights provides us with a list of internal capabilities and external conditions required for attaining a good life. An alternative to utilitarian measures of well-being, such as preference theories and hedonism, Nussbaum's account provides an objective but pluralistic account of what makes for a good human life. Susan Babbitt's (1996) work in political philosophy also argues for an objective account of the good.

What might a feminist theory of well-being look like? In general, do women lead better or worse lives than men? If women are happy with their lives, in unjust circumstances, does that count against subjective theories of well-being? These are important questions, and without a feminist account of the good—to guide lives and moral choices—we lack a feminist moral-theoretic account of ethics to answer them. It is my view that this area of feminist ethics, feminist theories of well-being, is a rich one for future feminist research. It is important that feminists engage in this area of philosophical work. It is not enough to leave developing plausible theories of the good to those who work only in consequentialist moral philosophy, and then reject consequentialism for its nonfeminist account of the good. Given that all reasonable moral theories require us (p. 629) to promote the good

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

some of the time, articulating what the good consists in is work best shared across approaches to ethics.

There are criticisms of a consequentialist approach to ethics other than the four I have discussed. I chose these points for discussion because I think they have often been presented with a feminist angle. But feminists sometimes raise other objections to consequentialism that are not grounded in feminist considerations. Consider the objection that consequentialism can be too demanding. This comes about because of the scope question, consequentialism's impartialist evaluation of outcomes, and the moral requirement that one maximize the good. An easy response to this worry is to pursue a nonmaximizing account of consequentialism, one in which consequences are still the factor that determines rightness of outcomes but also one according to which one need not maximize the good to do the right thing.

4. Feminist Objections to Consequentialism and Right Action

Our earlier feminist criticisms of consequentialism all focused on consequentialism and the good. Now assume that we can arrive at a feminist account of the good, either by defending a preference theory as distinct from a desire theory of well-being (Baber 2017), by revising desire theory as some have suggested (Driver 2005), or by moving to a feminist objective account of the good, such as perfectionism (Yuracko 2003). What about the consequentialist theory of right action? Are there feminist criticisms of consequentialism's account of the right? While these get less attention in the feminist ethics literature than criticisms of the theory of the good, because for the most part feminists and consequentialists agree about what the right actions are, I think that ultimately they pose a more serious challenge to the possibility of feminist consequentialism.

Consider the following example: A woman is brought to a police station and questioned about a crime. While there, she is stripped and searched. Unbeknownst to her, photos are taken and shared among the men who work in the station. No one knows her identity, and she never finds out about the photos. I ask my students if the police officers did anything wrong. They are convinced that a serious wrong has taken place even if there are no ill effects, even if the woman never finds out. What about the officers' pleasure at the taking, sharing, and viewing the photos? Does it count? My feminist ethics students might mull this briefly but say no. What if more people see the photos? What if they share them with police stations around the world? Is this getting better or worse? My students reject the idea that the consequences matter. Surely it cannot be getting better. This act is wrong, plain and simple. We know this without tallying up pleasures, they say. It's wrong because the woman's rights were violated. The results are morally irrelevant.

This is one example, but it is easy to come up with others. Just as some feminist intuitions line up with consequentialist moral reasoning—what do the results look like for (p. 630) the women involved?—other feminist intuitions are closer to Kantian deontology. Was

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

someone used merely as a means? Did this action treat someone as an object rather than a person?

If feminists are so critical of consequentialism, and if feminist moral thought is influenced by Kantian deontology (especially the concept of “respect”), one might wonder why there is so much feminist work on theories of the good? Given there is so much work on theories of the good, one might likewise wonder why there isn’t a burgeoning literature on feminist consequentialism. Julia Driver (2005) is often the lone voice articulating an explicitly full-blown feminist version of consequentialism. Yet there is a lot of feminist work on desire theory, adaptive preferences (Khader 2009; 2011; Superson 2005; Walsh 2015; Terlazzo 2016), and other theories of the good (Yuracko 2003).

I have three different answers to that question.

First, as I’ve argued elsewhere, feminist moral philosophy doesn’t draw a sharp line between the right and the good (Brennan 2005). Consequentialists and deontologists alike share the idea that considerations of the good are separate from accounts of right action. But why? I have called this shared assumption “plug and play ethics.” The idea is that we can take any account of good and plug it into any account of right action. This is not obviously true though. While pleasure might fit well with a maximizing account of right action—it’s true that more is better—other kinds of goods such as having a good character might fit best with a sufficiency account.

Second, the inclusion of constraints and permissions in a moral theory does not mean that there are no obligations to promote the good. It simply means that promoting the good is constrained by certain rights or rules. In the absence of those rights and rules, one may well be obligated to bring the good about.

Third, feminist work in ethics runs into feminist political philosophy in ways that make the line between the right and the good blurry (Calhoun 2005). Considerations of the good may be important in public life even if they don’t dominate personal ethics (Goodin 1995).

So all things considered, is there room for a feminist version of consequentialism? Julia Driver thinks so, arguing that consequentialism can accommodate feminist aims because it is responsive to empirical information, can accommodate the value of relationships in good lives, and is appreciative of distinctive vulnerabilities (Driver 2005).

But if we take the need to include rights or constraints seriously, is the theory that’s left consequentialist? This is partly a question of taxonomy in contemporary ethics. Shelly Kagan (1997), for example, counts any theory that includes rights or constraints as nonconsequentialist. Constraints are, on this common way of categorizing moral theories, the feature that makes a theory nonconsequentialist. A moderate deontological theory is one which includes overridable rights or constraints. In a moderate deontological theory, consequences matter. However, to count as consequentialist on this way of dividing up the

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ethical landscape, a theory must reject rights and constraints. But why divide up moral theories in this way? It makes consequentialism the more extreme view.

A moderate theory might well be a moderate *consequentialist* theory which gives significant scope—especially in public life, especially when no rights are involved—to (p. 631) promoting the overall good. I am more inclined to agree with Douglas Portmore (Chapter 1, this volume) that a consequentialist theory can include constraints. I think that if a theory gives considerable weight to consequences, then we ought to think of that theory as broadly consequentialist. In that sense, feminist moral theorizing is and always has been consequentialist.

What would make such a theory feminist?

Alison Jaggar argues that to count as feminist a moral theory must state (1) that the subordination of women is morally wrong and (2) that the moral experience of women is worthy of respect (Jaggar 1991, 95). Later, Jaggar phrases the second assumption differently, saying that "the moral experience of women should be treated as respectfully as the moral experience of men" (Jaggar 1991, 97–98). Elsewhere, I've argued that feminist ethical theories are those ethical theories that share two central aims: (a) to achieve a theoretical understanding of women's oppression with the purpose of providing a route to ending women's oppression and (b) to develop an account of morality which is based on women's moral experience(s). The first aim is normative (call this the "feminist conclusion requirement") and the second descriptive (call this the "women's experience requirement") (Brennan 1999). Finally, in her recent *Stanford Encyclopedia* entry on feminist ethics, Kate Norlock (2019) describes feminist ethics as a way of doing ethics, rather than a branch of ethics. That is, a feminist approach to ethics can be consequentialist, deontological, or virtue theoretic, but what makes it feminist is the approach to moral theorizing.

Given these desiderata, there is nothing that stops a consequentialist approach to ethics, suitably revised and rebuilt, as counting as a feminist approach to ethics. Indeed, I would argue more strongly that all feminist approaches to ethics ought to accord some role to consequences and results, and therefore ought to contribute to debates and discussions within consequentialist ethics.

References

Babbitt, S. 1996. "Transformation Experiences and Rational Deliberation." In *Impossible Dreams: Rationality, Integrity, and Moral Imagination*, 37–60. Boulder, CO: Avalon.

Baber, H. E. 2017. "Is Utilitarianism Bad for Women?" *Feminist Philosophy Quarterly* 3, no. 4, Article 6. doi:10.5206/fpq/2017.4.6

Baier, A. 1985. "What Do Women Want in a Moral Theory?" *Noûs* 19, no. 1: 55–63.

Bartky, S. L. 1990. *Femininity and Domination: Studies in the Phenomenology of Oppression*. New York: Routledge.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Bentham, J., and Crompton, L. (1785) 1978a. “Offences against One’s Self: Paederasty (Part 1). *Journal of Homosexuality* 3, no. 4: 389–405. doi:10.1300/J082v03n04_07

Bentham, J., and Crompton, L. 1978b. “Jeremy Bentham’s Essay on “Paederasty” (Part 2). *Journal of Homosexuality* 4, no. 1: 91–107. doi:10.1300/J082v04n01_07

Brennan, S. 1999. “Recent Work in Feminist Ethics.” *Ethics* 109, no. 4: 858–893. doi: 10.1086/233951

Brennan, S. 2005. “Review of *Setting the Moral Compass: Essays by Women Moral Philosophers*, C. Calhoun (Ed.).” *Ethics* 116, no. 1: 219–222. doi:10.1086/454375

Burgess-Jackson, K. 1995. “John Stuart Mill, Radical Feminist.” *Social Theory and Practice* 21, no. 3: 369–396. **www.jstor.org/stable/23557193**.

(p. 632) Calhoun, C. 2005. *Setting the Moral Compass: Essays by Women Philosophers*. Oxford: Oxford University Press.

Card, C. 1996. *The Unnatural Lottery: Character and Moral Luck*. Philadelphia: Temple University Press.

Dea, S. 2016. “A Harm Reduction Approach to Abortion.” In *Without Apology: Writings on Abortion in Canada*, edited by S. Stettner, 317–332. Edmonton: Athabasca University Press.

Driver, J. 2005. “Consequentialism and Feminist Ethics.” *Hypatia* 20, no. 4: 183–199. doi: 10.1111/j.1527-2001.2005.tb00543.x

Duflo, E. 2012. “Women Empowerment and Economic Development.” *Journal of Economic Literature* 50, no. 4: 1051–1079. doi:10.1257/jel.50.4.1051

Flanagan, J., and Watson, L. 2019. *Debating Sex Work*. Oxford: Oxford University Press.

Goldstein, L. F. 1980. “Mill, Marx, and Women’s Liberation.” *Journal of the History of Philosophy* 18, no. 3: 319–334. doi:10.1353/hph.2008.0726

Goodin, R. E. 1995. *Utilitarianism as a Public Philosophy*. Cambridge Studies in Philosophy and Public Policy. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511625053

Held, V. 1990. “Feminist Transformations of Moral Theory.” *Philosophy and Phenomenological Research* 50:321–344. doi:10.2307/2108046

Jaggar, A. 1991. “Feminist Ethics: Projects, Problems, Prospects.” In *Feminist Ethics*, edited by C. Card, 78–106. Lawrence: University of Kansas Press.

Kagan, S. 1997. *Normative Ethics*. Boulder, CO: Westview Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Khader, S. J. 2009. “Adaptive Preferences and Procedural Autonomy.” *Journal of Human Development and Capabilities* 10, no. 2: 169–187. doi:10.1080/19452820902940851

Khader, S. J. 2011. *Adaptive Preferences and Women’s Empowerment*. Oxford: Oxford University Press.

Kuhse, H., Singer, P., and Rickard, M. 1998. “Reconciling Impartial Morality and a Feminist Ethic of Care.” *Journal of Value Inquiry* 32, no. 4: 451–463. doi:10.1023/A:1004327810964

Kymlicka, W. 2001. *Contemporary Political Philosophy: An Introduction*. Oxford: Oxford University Press.

Mackenzie, C., and Stoljar, N., eds. 2000. *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*. New York: Oxford University Press.

McLeod, C. 2002. *Self-Trust and Reproductive Autonomy*. Cambridge, MA: MIT Press.

Mill, J. S. (1869) 1984. *The Subjection of Women*. In Vol. 21 of *Collected Works of John Stuart Mill*, edited by J. M. Robson. Toronto: University of Toronto Press.

Nagel, T. 1986. *The View from Nowhere*. Oxford: Oxford University Press.

Noddings, N. 1984. *Caring*. Berkeley: California University Press.

Norlock, K. 2019. “Feminist Ethics.” In *The Stanford Encyclopedia of Philosophy* (Summer 2019 edition), edited by E. N. Zalta. **https://plato.stanford.edu/archives/sum2019/entries/feminism-ethics/**.

Nozick, R. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.

Nussbaum, M. 1995. “Human Capabilities, Female Human Beings.” In *Women, Culture and Development: A Study of Human Capabilities*, edited by M. Nussbaum and J. Glover, 61–104. Toronto: Oxford University Press.

Nussbaum, M. 1999. *Sex and Social Justice*. New York: Oxford University Press.

Nussbaum, M. 2000. *Women and Human Development: The Capabilities Approach*. Cambridge: Cambridge University Press.

Okin, S. M., and Mansbridge, J. 2005. “Feminism.” In *A Companion to Contemporary Political Philosophy*, edited by R. E. Goodin and P. Petit, 269–290. Oxford: Blackwell.

(p. 633) Sen, A. 1990. “More Than 100 Million Women Are Missing.” *New York Review of Books* 37, no. 20.

Sen, A. 2000. *Development as Freedom*. New York: Anchor Books.

Sherwin, S. 1984. “A Feminist Approach to Ethics.” *Dalhousie Review* 64, no. 4: 704–713.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Sherwin, S. 1991. “Abortion through a Feminist Ethics Lens.” *Dialogue* 30, no. 3: 327–342. doi:10.1017/S0012217300011690

Singer, P. 1993. *How Are We to Live? Ethics in an Age of Self-Interest*. Melbourne: Text Publishing.

Superson, A. 2005. “Deformed Desires and Informed Desire Tests.” *Hypatia* 20, no. 4: 109–126.

Temkin, L. 1986. “Inequality.” *Philosophy & Public Affairs* 15, no. 2: 99–121. **https://www.jstor.org/stable/2265381**

Terlazzo, R. 2016. “Conceptualizing Adaptive Preferences Respectfully: An Indirectly Substantive Account.” *Journal of Political Philosophy* 24, no. 2: 206–226. doi:10.1111/jopp.12062

Tessman, L. 2017. *When Doing the Right Thing Is Impossible*. New York: Oxford University Press.

Thomson, J. J. 1985. “The Trolley Problem.” *Yale Law Journal* 94, no. 6: 1395–1415. doi:10.2307/796133

Walsh, M. B. 2015. “Feminism, Adaptive Preferences, and Social Contract Theory.” *Hypatia* 30, no. 4: 829–845. doi:10.1111/hypa.12175

Williams, B. 1973. “A Critique of Utilitarianism.” In *Utilitarianism: For and against*, edited by J. J. C. Smart and B. Williams, 77–150. Cambridge: Cambridge University Press.

Williams, B. 1976. “Persons, Character, and Morality.” In *The Identity of Persons*, edited by A. O. Rorty, 197–216. Berkeley: University of California Press.

Yuracko, K. A. 2003. *Perfectionism and Contemporary Feminist Values*. Bloomington: Indiana University Press.

Notes:

([1]) I know that traditionally we have defined consequentialism as holding that bringing about a world that’s better overall is the *only thing* that matters, but that strikes me as an unnecessarily restrictive way to think about consequentialism. Imagine a view that required us to bring a better world to some minimal standard. That view on the traditional definition of consequentialism would count as consequentialist since all that matters is consequences. But what about a view according to which the right action is the one that maximizes the good unless doing so would violate a rights claim? If there were very few rights, then the second view would be more demanding, in terms of bringing about the good, but not consequentialist. In fact, I think the narrow definition of consequentialism—according to which if we’re ever not required or not permitted to bring about the good, the theory counts as nonconsequentialist—is responsible for moral theorists of various

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

stripes, including feminist moral theorists, not recognizing the consequentialist elements of their views.

([2]) Jeremy Bentham, “Offences Against One’s Self” (see Bentham and Crompton 1978a; 1978b).

([3]) “The question is not can they reason? Nor, can they talk? But can they suffer?” (J. Bentham, *An Introduction to the Principles of Morals and Legislation* (1789), chap. xvii.)

([4]) Jeremy Bentham, “Proposal for a New and Less Expensive Mode of Employing and Reforming Convicts” (London, 1798).

([5]) See Jessica Flanagan’s contribution to *Debating Sex Work*, by Jessica Flanigan and Lori Watson (Oxford University Press, 2019).

([6]) See Dea (2016).

([7]) Is Mill a radical feminist or a liberal feminist? See Goldstein (1980) and Burgess-Jackson (1995).

([8]) Whether or not *The Subjection of Women* was coauthored, Harriet Taylor Mill had covered some of the same arguments in her 1851 article “The Enfranchisement of Women,” published anonymously in the *Benthamite Westminster Review.*

([9]) See Thomson (1985).

([10]) For a description of the role equality plays in utilitarian political thought, see Kymlicka (2001).

([11]) Thanks to Douglas Portmore for this way of expressing the worry.

([12]) This is captured very well in the title of Thomas Nagel’s *The View from Nowhere*.

([13]) See Bernard Williams (1973; 1976).

Samantha Brennan

Samantha Brennan is Dean of the College of Arts and Professor of Philosophy at the University of Guelph. Her research focuses on contemporary normative ethics, including feminist ethics. A recent area of focus for her work is children’s rights, parents’ rights, and issues of family justice. She’s also written and published about micro-inequities, the climate issue in philosophy departments, the moral significance of fashion, and the badness of death.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Consequentialism and Reasons for Action

Christopher Woodard

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.31

Abstract and Keywords

Consequentialist theories often neglect reasons for action. They offer theories of the rightness or the goodness of actions, or of virtue, but they typically do not include theories of reasons. However, consequentialists can give plausible accounts of reasons. This chapter examines some different ways in which such accounts might be developed, focusing on act consequentialism and rule consequentialism and on the relationship between reasons and rightness. It notes that adding claims about reasons to consequentialist theories may introduce a welcome kind of complexity, and in doing so it may help to make consequentialist approaches to ethics more appealing. For example, it may help consequentialists to explain the ideas of moral constraints and moral options.

Keywords: reasons, act consequentialism, rule consequentialism, rightness, moral constraints,, moral options

DISCUSSION of reasons for action is widespread in moral philosophy and in everyday ethical thought and talk. Reflecting on instances where someone has done something wrong, we might note that there was at least some reason for what she did. Wondering what to do, we may try to weigh up the reasons for and against some action. Yet consequentialist theories typically neglect reasons, focusing only on the rightness of actions. Or, if other topics are introduced, this is often a matter of extending consequentialist evaluation to other items, such as character traits or policies or institutions. With some notable exceptions, consequentialists often do not make explicit theoretical claims about reasons for action (exceptions include Crisp 2006, chap. 2; McElwee 2010; and Portmore 2011). Since reasons for action are among the things we care about, and would like to understand better, it is worth considering what consequentialist approaches to ethics can say about them.

As I shall understand it, “consequentialism” is an approach to ethics which seeks to explain matters of ethical interest (such as the rightness of actions, or the justice of institutions, or reasons for action) in terms of the goodness of outcomes. In discussions of consequentialism, “outcome” is understood in an especially broad way, such that “the outcome of X” includes everything that would happen if X were realized, including the real-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ization of X itself. So, for example, the outcome of an action includes the performance of the action (Portmore 2011, 57). We should also understand "goodness" broadly, not building in more than is necessary to the definition of "consequentialism." Thus, let us say that consequentialism may employ an agent-neutral or an agent-relative account of goodness, and that it may understand goodness either as prior to reasons, or as a matter of what we have reason to desire (for discussion see Louise 2004; Schroeder 2007; and Portmore 2011, chap. 3). Finally, note that consequentialism as I have defined it does not include a commitment to direct evaluation of anything. It allows for the possibility (p. 180) that what makes X favored (right, just, virtuous, or legitimate, say) is the value of Y (Kagan 2000, 134–155).[1]

Let us begin by considering the concept of normative reasons for action, before exploring what act consequentialists might say about reasons, and then exploring what indirect forms of consequentialism might say about them. My aim will be to describe some of the theoretical options open to consequentialists who wish to give theories of reasons and thereby to contribute to further exploration of those options.

1. Reasons for Action

It is not possible to give an entirely uncontroversial characterization of the concept of normative reasons for action. I will not try to survey all of the controversies here.[2] Very broadly, I take the concept of a normative reason for action to be the concept of a consideration in favor of or against acting in some way.[3] This is to be contrasted, on one hand, with the concept of motivating reasons, which is the concept of a consideration that moves an agent to act by making it seem to her to be a good thing to do.[4] On the other hand, we can distinguish normative reasons for action from other normative reasons: there may be considerations in favor of or against believing something, desiring something, or hoping for something, for example.

One central controversy about normative reasons for action—"reasons for action" for short, from now on—is whether an agent's perspective in some way constrains what reasons she has. This controversy is connected to background views about the relationships between the concept of reasons for action and other concepts. For example, we might think that there is a close connection between the concept of reasons for action and the concepts of good deliberation, or of blameworthiness. If so, we are likely to think that an agent's perspective in some way constrains what reasons for action she has, since good deliberation must connect in some way with the agent's perspective, and how things seemed to an agent is relevant to her blameworthiness. Alternatively, we might think that the concept of reasons for action is more closely connected to the concept of the rightness of actions than to good deliberation or blameworthiness. If so, the issue about perspective is left more open. On one view about rightness, the rightness of actions is not at all constrained by the agent's perspective. Something could then be a reason in (p. 181) favor of an agent acting in some way whether or not she could possibly be aware of it, given her situation and her perspective.[5]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

If we emphasize the connection with good deliberation or blameworthiness, we are bound to think of reasons as in some way perspectival; if we emphasize the connection with rightness, we can be more open-minded about whether they are perspectival, since we may or may not think that rightness itself is perspectival. In this chapter I will assume that there is a close connection between reasons and rightness, leaving aside the question of the connections between reasons, good deliberation, and blameworthiness.[6] This enables us also to set aside the issue of whether an agent's reasons are constrained by her perspective.

Consider, then, the relationship between the concept of reasons for action and the concept of rightness of actions. A reason is a consideration in favor of, or against, performing some action. An action is right if and only if it is not wrong to perform it.[7] Intuitively, if there is a reason in favor of performing some action, that is certainly relevant to whether it is right to perform it. However, it does not seem to entail that it is right to perform it. For example, it may be wrong on some occasion to lie to a friend—even though doing so would protect his feelings, and this is a reason in favor of lying to him. Reasons for action may conflict with each other, so that there is a reason in favor of acting in some way but a stronger reason against acting in that way. Moreover, there may be more than one reason for or against any action. This suggests that the relationship we are after is between the rightness or wrongness of an action and the *overall set* of reasons for or against it.

Can we say more about this relationship? I shall assume, more specifically and controversially, that if there is *sufficient reason* to perform an action, then it is *not wrong* to perform it (which is to say that it is right, i.e., either required or optional). This leaves open a number of issues. First, it leaves open whether an action must be wrong when there is not sufficient reason to perform it. Perhaps it need not be: for example, perhaps there is no reason, and so not sufficient reason, to take *this* can of soup rather than its identical neighbor from the supermarket shelf, even though it is not wrong to take it.[8] Second, it leaves open what counts as "sufficient reason." One possibility is that there is sufficient reason to perform an action if and only if there is no weightier reason (or combination of reasons) in favor of any alternative. But, as we shall see later, that is not the only possibility. Third, it leaves open what explains the correlation between rightness and presence of sufficient reason. This could be because presence of sufficient reason (p. 182) makes it true that the action is right; alternatively, it might be that the balance of reasons, and the rightness of actions, are each to be explained in terms of the goodness of the outcome of the action.

Finally, the assumption that an action is right whenever there is sufficient reason to perform it leaves open the domain of reasons and the corresponding sense of rightness and wrongness. It is common to distinguish between kinds of normative reason for action. For example, we might distinguish moral reasons from legal or prudential reasons. We might also say that a certain action was legally right but morally wrong, for example. In the case of moral rightness, an interesting further issue is whether the relevant domain of reasons is all reasons, or only moral reasons. Could there be most reason to do what is morally wrong?[9] In what follows I will leave these interesting issues about kinds of rea-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

sons aside, conducting the discussion instead in terms of a generic concept of normative reasons for action, and a sense of rightness that is left open deliberately. In this generic sense, an action is right whenever there is sufficient reason to perform it.

A reason for action, then, is a consideration in favor of or against performing some action. Which reasons an agent has may or may not be constrained in some way by her perspective. Her reasons may conflict with each other, and so she may have reasons to perform actions that it would be wrong to perform. But when she acts in ways for which she has sufficient reason, she does not act wrongly.

2. Act Consequentialism

Consider a simple form of act consequentialism, which says that an action is right if and only if there is no relevant alternative action with a better outcome. That is, this theory claims that there is a strict correlation between the rightness of actions and the goodness overall of their outcomes.

Reasons for action, we have assumed, may conflict with each other—so that there may be one or more reasons to do something that is wrong. There may also be more than one reason to do something that is right. The rightness of actions is related to the overall set of reasons, we assumed. According to act consequentialism, it is also related to the goodness overall of their consequences. If act consequentialism (and our assumption) is true, then, the rightness of actions is related both to reasons overall and to goodness overall.

This suggests a natural hypothesis: for act consequentialists, reasons are to be explained in terms of the different respects in which outcomes are good. Just as there may be a reason in favor of a wrong action, it is also the case that a wrong action may have an outcome that is good in some respect, even though it is not best overall. The contrast between goodness in some respect and goodness overall is a natural way for act (p. 183) consequentialists to relate the idea of there being a reason in favor of an action to the idea that it is right. According to this hypothesis, for act consequentialists rightness and wrongness correlate with goodness overall, while individual reasons correlate with individual respects in which an outcome may be good. To return to our previous example: lying to your friend has an outcome which is good in one respect, since his feelings are not hurt; but if lying is nevertheless wrong, this outcome must be worse overall than the outcome of not lying to him.

On the definitions we have given, act consequentialism is compatible with a very wide variety of claims about which features of outcomes may be good in some respect.[10] This means that act consequentialists may claim, for example, that each token action of keeping a promise is in itself good. Thus they may claim that there is always a reason to keep promises. Or, to take a different example, they may claim that it is good that people get what they deserve, and thus that there is always a reason to give people what they deserve.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The hypothesis looks promising, but we must take care in spelling it out further. The difficulty is to spell out exactly what must be true of a feature of an outcome for its value to supply a reason to produce it, according to act consequentialism. One implausible answer to this question is that it is sufficient that the feature has positive value. This cannot be right. Suppose that everyone is currently blissful. You could either do nothing, in which case everyone's bliss would remain undisturbed, or you could press a switch, in which case everyone would be reduced to whichever level of well-being is minimally positive. It does not suffice for you to have a reason to press the switch that it would produce an outcome in which each person's well-being is positive. Intuitively, you have no reason at all to press the switch.

It is more promising to claim that a feature of an outcome supplies a reason to produce it if and only if it is better than *the corresponding feature* of some *relevant alternative outcome*.[11] In the case just examined, it is plausible to claim that the positive features of the outcome of pressing the switch—the barely positive level of well-being of each person—are all worse than the corresponding features of the relevant alternative outcome, where that means each person's level of well-being if you do not press the switch. If that is right, then this proposal correctly implies that you have no reason to press the switch, even though it results in an outcome with valuable features.

However, it is not trivial to provide a satisfactory general account of either of the elements of this proposal—that is, the idea of a "corresponding feature" and the idea of a "relevant alternative outcome." One difficulty attending the idea of a corresponding feature arises in "nonidentity" cases. Suppose that you must do either A or B. If you do A, Tom will be born and will have a good life. If you do B, Tim will be born instead of (p. 184) Tom and will have an excellent life. Do you have a reason to do A? Tom's good life is a good feature of the outcome of doing A—but is it better than the corresponding feature of the outcome of doing B? That depends on whether we take the corresponding feature to be *Tom's* well-being in the world in which you do B, or instead to be *Tim's* well-being in that world. If a good life is better than no life, you have a reason to do A if we say that the corresponding feature is Tom's well-being, but not if we say that it is Tim's well-being (Parfit 1987, chap. 16; Roberts and Wasserman 2009).

What about "relevant alternative outcome"? Act consequentialists should probably say that the relevant outcomes are the outcomes of the agent's other options, where an option is an alternative action she could perform in the circumstances (see Smith, Chapter 6, this volume; and Portmore 2019). This seems to get many cases right. Suppose that Sylvie and Shreya are both badly off, and that you could either make Sylvie somewhat happy or Shreya very happy. The best outcome overall is produced by making Shreya very happy, let us say, so according to act consequentialism, that is the right thing to do. Nevertheless, intuitively there is some reason to make Sylvie happy. Her happiness is a consideration in favor of doing that. If you were to make her happy, you would have acted wrongly according to act consequentialism, but there would have been a reason for acting in that way (for related discussion, see Chappell 2015). Now, if we specify the "relevant alternative outcome" as the outcome of your other option, we can explain this rea-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

son. Sylvie's happiness is good compared with her state if you benefit Shreya instead. So on the current proposal Sylvie's happiness supplies a reason to benefit her.

However, specifying relevant alternative outcomes in this way may seem to yield the wrong implications in other cases. Sometimes it is tempting to say that the relevant comparison is with the state of the world prior to action. Suppose that Lulu and Lisa are currently miserable. Suppose further that you could either make just Lulu happy, or both Lulu and Lisa happy, and that these are your only two options. Arguably, you have some reason to make just Lulu happy, even though it would be wrong to do so according to act consequentialism. Her happiness seems to be a consideration in favor of making her happy. But we cannot explain this if we compare her happiness with her state if you were to take your other option, since she is equally happy in that outcome. Thus, we might be tempted to take Lulu's miserable state before you acted to be the relevant comparison.

Though it might seem intuitively correct in this case, act consequentialists should probably resist the temptation to specify the relevant alternative outcome in this way. One reason is that in other cases the comparison with the state of the world before the agent acts seems to give clearly the wrong answer (Norcross 1997, 8–9). Another, more theoretical, reason is that it would mean that the act consequentialist account of rightness of actions (which *is* standardly defined in terms of the agent's options) would fall out of step with its account of reasons, since they would be based on comparisons with different sets of alternatives.

So we have the outline of an act consequentialist theory of reasons, but it leaves open some issues. The outline account is this: an agent S has a reason to perform some action A in circumstances C if and only if and because S could do A in C and the outcome of A in C would have some feature F which is better than the corresponding feature of a relevant (p. 185) alternative outcome.[12] As we have just noted, there are some difficulties in specifying what is to count as a corresponding feature, and which are the relevant alternative outcomes. In addition, of course, an act consequentialist must also give some account of goodness if she is to reach determinate conclusions about which reasons there are.

If an act consequentialist can answer these questions, she can produce a theory of reasons for action. The theory would tell us which reasons an agent has in any specified circumstances, and why. Since act consequentialism already contains a theory of rightness of actions, if we also assume that an action is right whenever there is sufficient reason to perform it, she will also have specified the notion of sufficient reason. There is sufficient reason to perform an action, according to act consequentialism, if and only if there is no relevant alternative action with a better outcome. The most natural interpretation of this claim is as reflecting two underlying thoughts. One is that the strength of a reason is proportional to the value of the feature that supplies it, and the other is that there is sufficient reason to perform an action only if it is favored by the strongest reasons overall. If we make these further claims, we get the correspondence between best overall outcomes and rightness of actions that act consequentialism asserts.[13]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

3. Indirect Consequentialism

Now consider a standard form of rule consequentialism, according to which an action is right if and only if it is permitted by the best set of rules. Whereas act consequentialism asserts a correlation between the rightness of actions and the goodness overall of their outcomes, rule consequentialism asserts a correlation between the rightness of actions and the goodness of sets of rules. The specific correlation it asserts is that right actions are in every case permitted by the best set of rules, and that wrong actions are, in every case, prohibited by the best set of rules.

There are important issues arising from the question of how to evaluate sets of rules. Since rules do not, all by themselves, have causal consequences, they must be "embedded" in some way, which means drawing some appropriate connection between the rules and things with causal consequences (Kagan 2000). In brief, there are two main issues. One is whether we should characterize the consequences of a set of rules in terms of the consequences of compliance with those rules, or instead in terms of the consequences of (p. 186) acceptance of those rules. Neither compliance nor acceptance entails the other, so we would expect different results depending on which way we go on this issue (Hooker 2000, 75–80). The other issue is whom we take to comply with, or accept, the rules in order to characterize the consequences of the rules. There are many possible answers to that question, and many have been explored in recent discussion (for example, see Ridge 2006; and Smith 2010). At one extreme, we could characterize the consequences of a set of rules in terms of the consequences of the agent alone complying with (or accepting) them on a single occasion. At the other extreme, we could characterize consequences in terms of everyone everywhere always complying with (or accepting) them. The first extreme option would result in a theory much like act consequentialism, while the second extreme option would result in something more similar to Kantianism. Many possible versions of rule consequentialism lie between these extremes (Woodard 2013).

In the case of act consequentialism, we explored the hypothesis that the fact that an outcome may have different good features explains how there can be multiple reasons, including reasons to perform wrong actions. Since rule consequentialism also appeals to the value of outcomes—albeit the outcomes of sets of rules—we could in principle explore the same hypothesis in relation to it. But a more obvious strategy is to try to explain the plurality of reasons for or against a single action in terms of the plurality of rules governing a single action. To return to our earlier example, we might say that there is a rule about lying, and a rule about looking after friends, which together explain why you have a reason to lie to your friend to protect his feelings, even though (we assumed) it would be wrong to do so. The rule governing lying is associated with a general reason not to lie, while the rule governing friendship is associated with a general reason to promote one's friends' well-being. (No doubt the best rules are more complex than this.) In this instance, these reasons conflict with each other, and it happens that the reason not to lie is stronger. It is natural to think of the rules that feature in the best set as associated with different considerations or reasons.[14]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Exactly how we should think of these rules depends on the way we answer the questions about embedding. If we formulate rule consequentialism in terms of compliance, the rules specify behavior. The consequence of complying with a rule that says “do not lie” is the consequence of people not lying. Which people? That is specified by our answer to the second embedding question. At one extreme, it is the agent alone on a single occasion. At the other, it is everyone at all times in all places. Rule consequentialists may, but need not, believe that the embedding questions must be answered realistically. They might take a deliberately idealizing approach, because they think that it is part of the concept of morality that it consists of rules that form an ideal code, in the sense that things would go well if (more or less) everyone lived up to them (Hooker 2000, 1 and 80–85).

Of course, we have not got a determinate outcome if we say only that it consists of people not lying. Not lying in which way? There are many different ways of not lying on each occasion on which to lie is an option, not to mention the many different ways of behaving when lying is not an option. Rule consequentialists must say more if they are to (p. 187) characterize the consequences of rules. One thing they say is that the consequences to consider are the consequences of the *whole set* of rules. So, if we are characterizing these in terms of compliance, the relevant outcome would be one in which the relevant agents comply with all of the rules at once. That is likely to be much more determinate. Indeed, we might then worry about the opposite problem, of it being impossible to comply with all of the rules at once. Rule consequentialists must either tailor the rules so that they do not conflict—or, more likely, address the issue of what the set of rules requires when it is not possible to comply with all members of the set at once. This is likely to involve the idea that some rules take precedence over others.

However all of these questions are settled, compliance with a set of rules is ultimately a pattern of behavior. To comply is to behave in a way that the rule permits, and so compliance overall is a concatenation of pieces of behavior, by one or more agents. Suppose, as we have been doing, that complying with the best set of rules involves you not lying to your friend, even at the cost of hurting his feelings. What account can the rule consequentialist give of your reason not to lie on this occasion?

The rule consequentialist might say, simply, that the source of the reason is the rule. But that invites worries about rule fetishism. If he wishes to go further, one thing he could say is that not lying on this occasion is *your part now* in the pattern of behavior that consists of all relevant agents complying with the best set of rules. This answer relies on a common idea, which I have elsewhere labeled the idea of “pattern-based reasons” (Woodard 2013; and Woodard 2019, chap. 5). The idea is that the fact that some action is part of a favored pattern of action can provide a reason to perform the action. In the current context, what makes the pattern “favored” is that it has better consequences than any alternative pattern performable by the same set of agents (namely, those specified in answer to the second embedding question). Assuming that this includes other agents, we could say that your not lying on this occasion is your part, right now, in the best that the whole set of agents could do. This is to treat the *parthood* relation between a token action and a favored pattern of action as providing a reason to perform the part. Schematically, the

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

idea is this: agent S has a pattern-based reason to perform action A in circumstances C if and only if and because S could do A in C and A is S's part, in C, of some favored and eligible pattern of action P.

The idea of pattern-based reasons appears to be part of common ethical thought. It is common for people to explain their actions in terms of larger patterns of action of which they are parts, and which they take to be good or right (for some evidence, see Bardsley et al. 2010). Moreover, it is common for people to use the language of parthood to explain their thinking: people often say "I want no part in that" or "you should play your part," for example. So, if rule consequentialists appeal to the idea of pattern-based reasons in their account of reasons for action, they are, at least, appealing to an idea with common currency. Ordinary ethical thought appears to recognize playing one's part as the source of a kind of reason, as well as recognizing the kind of reason that act consequentialism articulates, of causing good outcomes.

The idea of pattern-based reasons raises a number of theoretical puzzles, however. A central puzzle concerns the concept of "eligibility" of patterns. This marks the difference (p. 188) between those possible patterns of action which do, and those which do not, generate reasons to perform their parts. Presumably not every good possible pattern of action generates pattern-based reasons. What, then, distinguishes those which do (the eligible) from those which do not (the ineligible)? If we interpret rule consequentialism as employing the idea of pattern-based reasons, answers to the second embedding question (about which agents to specify, when characterizing the consequences of a set of rules) are, in effect, answering the question about eligibility. If we say that the consequences of the set of rules are the consequences of the agent on this occasion complying with those rules, we are in effect saying that the only eligible pattern is one consisting of this agent's action on this occasion.[15] If, on the other hand, we say that the consequences of the set of rules are the consequences of every agent on every occasion complying with those rules, we are in effect saying that this highly extended, highly idealizing pattern of action is eligible. One question for those interested in this sort of theory is thus whether we can give satisfying explanations of why some patterns should be treated as eligible and others not.

A related question is whether pattern-based reasons, if they exist, would have any practical significance. This is related to the "collapse worry" about rule consequentialism (Hooker 2000, chap. 4). In the current context, the issue is whether pattern-based reasons would make any difference to the total set of reasons agents have, or to which actions are right. The answer depends on how we specify eligibility. It is pretty clear that my part in the best that every agent could do might be different from the best thing I can do taking as given others' nonideal behavior. On the other hand, an idealizing answer to the question about eligibility might be thought to purchase practical significance at the cost of plausibility. Critics of idealizing theories ask why we should think that the fact that some pattern would have good consequences can provide a reason to play one's part in it, if this pattern would not be realized because others would not play their parts (Dietz 2016; Parfit 2011, 312–320; Podgorski 2018). So an issue for rule consequentialists, on

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

this way of understanding their view, is whether it is possible to specify eligibility in a way that is both plausible on general grounds and yields practical significance.

A further puzzle is how to recover the idea of plural, possibly conflicting, reasons if we appeal to the consequences of a whole set of rules. The consequences of compliance with a set of rules are the consequences of a pattern of behavior. But how do we get plural reasons from a single pattern? When you do not lie to your friend, you are (we have been assuming) playing your part in the pattern of behavior that constitutes compliance (by the relevant agents) with the best set of rules. So we can use the idea of pattern-based reasons to offer an explanation of your reason not to lie. But what can we say about the other reason we supposed you have, to lie so as to protect your friend's feelings?

(p. 189) One option would be to treat this as a reason of the sort that act consequentialists recognize. Lying on this occasion would protect your friend's feelings, and that is a good feature of this outcome. This would be to think of rule consequentialism as a pluralist theory, recognizing both pattern-based reasons and the kind of reasons that act consequentialists recognize. This is, on independent grounds, a plausible interpretation of rule consequentialism, at least when it is specified in terms of compliance. For rule consequentialists tend to claim that the best set of rules includes rules instructing agents to do the best they can, as individuals, in the circumstances. One example of this is the "disaster prevention rule" that rule consequentialists tend to recognize (Brandt 1992, 151; Hooker 2000, 98–99).[16] This instructs agents to break any other rule when necessary to prevent a disaster. But this is not plausibly interpreted as a pattern-based reason: it is the good outcome of breaking the rule on this occasion that generates the reason, not the good outcome of a larger pattern of rule breaking.

So rule consequentialists might explain plurality simply by postulating pattern-based reasons associated with a single pattern (which reasons could not conflict among themselves) together with the reasons that act consequentialists recognize (with which the pattern-based reasons could conflict). But another possibility would be to evaluate the contribution of individual rules to the consequences of the best set. The obvious way to do this would be to compare the consequences of the best set with a series of alternative sets, each of which differs from the best set by removing a single rule at a time. Thus, for example, we get a sense of the contribution of the best rule governing lying by considering the difference between the consequences of compliance with the best set B, and the consequences of compliance with a different set of rules B* which lacks only that rule. We can then think of the rule governing lying as associated with a general reason against lying, which is a pattern-based reason to play one's part in the pattern specified by B rather than the pattern specified by B*.[17]

Finally, we should return to the first question about embedding—which was whether to characterize the consequences of a set of rules in terms of compliance with them or acceptance of them. The discussion so far has assumed that we go with compliance. This enables us to think of the consequences of a set of rules as being the consequences of a pattern of behavior, and thereby to employ the idea of pattern-based reasons. If instead

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

we go with acceptance, we might have to give a different account of rule consequentialist reasons. As Hooker understands acceptance, for example, "to accept a code of rules is … to have *a moral conscience of a certain shape*. In other words, when rule-consequentialists consider alternative codes of rules, they are considering alternative possible contours for people's consciences" (Hooker 2000, 91, emphasis in original). As this suggests, rule consequentialism of this sort is similar to motive consequentialism, which is another indirect form of consequentialism. According to a standard formulation (p. 190) of motive consequentialism, an action is right if and only if it would be performed in the circumstances by an agent with the best motives (Parfit 2011, 375).[18]

Can we furnish Hooker-style rule consequentialism, or motive consequentialism, with a theory of reasons along the same lines as the one we considered for compliance-style rule consequentialism? That depends on whether the consequences of a moral conscience, or a set of motives, can be cashed out entirely in terms of a pattern of behavior. Such a pattern would not consist of compliance with the best set of rules, but of behavior generated by the best set of motives or the best conscience. This behavior is likely to include the behavior of people other than the agent with the motive or character. For example, if someone is highly disposed to break moral rules, this disposition might cause others to behave warily around him. If the consequences of the best conscience or set of motives can be cashed out in terms of behavior in this way, then we could apply the concept of pattern-based reasons in roughly the same way as we did with compliance-style rule consequentialism (albeit with reference to different patterns of behavior, which bear a different relation to the rules constituting the best set). If, on the other hand, the consequences of motives and consciences extend beyond the consequences of behavior, these forms of indirect consequentialism would have to find some alternative way of explaining agents' reasons for action.

4. Consequentialism and Constraints

One important issue for consequentialist theories of reasons is whether they can account for all of the reasons that we believe exist. An interesting instance of that question is whether consequentialist theories of reasons can explain the existence of moral constraints.

The idea of moral constraints is that it is sometimes morally wrong to act in a way that has the best outcome overall, impersonally evaluated. This idea seems to be one component of the idea of moral rights. If Smith has a moral right to bodily integrity, it may be wrong for a beneficent surgeon to seize his organs to save five others, even when seizing them would make the outcome best overall, impersonally evaluated. Smith's moral right acts as a constraint in the sense that it constrains what may be done permissibly to promote the impersonal good (Kagan 1989, 4, 24–32; Scheffler 1982).[19]

One natural way to interpret the idea of moral constraints is to think of them as applying to *kinds* of action. For example, we might think of the kind of action *seizing someone's organs* as governed by a constraint, which applies in the case just mentioned. We can then

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

distinguish between absolute and nonabsolute moral constraints. An absolute moral constraint implies that every token action of the proscribed kind is morally (p. 191) wrong, while a nonabsolute moral constraint fails to imply this. A nonabsolute moral constraint might imply instead that there is a reason—perhaps a strong reason—not to perform actions of that kind (Kagan 1989, 4–5). Note also that constraints need not be, in an intuitive sense, "negative" in the way that the constraint against seizing organs is negative. There could be a constraint requiring gratitude in response to kindness, for example. Such a constraint would imply that it is sometimes wrong not to express gratitude, even when necessary to make the outcome best overall, impersonally evaluated. Intuitively, we might describe this by saying that we are constrained to express gratitude (though we could equally describe this negatively, of course, by saying that there is a constraint against failing to express gratitude).

The idea of constraints is very important in many ethical views. As already noted, it is one component of the idea of moral rights. It is also one component of standard views of the moral force (in some circumstances) of legal rights. If you have a legal ownership of your driveway, it is at least sometimes wrong for others to park on it without your permission, even if doing so has the best outcome.[20] More broadly, many ethical views attribute significance to kinds of action in a way that the idea of constraints seems to capture. Ross's objection to act utilitarianism, for example, was that it (wrongly) implies that it is right to break a promise whenever doing so makes the outcome even a whisker better. A natural way of interpreting this objection is to think of Ross as claiming that there is a reason to keep promises (a reason against breaking them) because of the kind of action that *keeping promises* is—a reason which the utilitarian fails to take into account (Ross 1930/2002, 34–35).

Of course, one option for consequentialists is to deny the existence of moral constraints. But it is worth considering whether their theories of reasons could accommodate them. Consider first whether act consequentialism can do so. Recall that, according to our outline account, act consequentialism claims that an agent S has a reason to perform some action A in circumstances C if and only if and because S could do A in C and the outcome of A in C would have some feature F which is better than the corresponding feature of a relevant alternative outcome. Let us suppose that action A is of the kind *keeping promises*. Act consequentialism is compatible with very many theories of the goodness of outcomes, including those that claim that it is noninstrumentally good to keep promises. So, when combined with such a theory of goodness, act consequentialism will imply that there is always a reason to keep a promise, in virtue of the fact that keeping it will have an outcome with a feature (the fact that the promise is kept) which is better than the corresponding feature (the fact that the promise was broken) of a relevant alternative outcome (the outcome of breaking the promise).

This goes some way toward accounting for a constraint against breaking promises, but it is not yet enough. Act consequentialism claims that an action is right if and only if there is no relevant alternative action with a better outcome. The idea of a constraint, we said, was that it is sometimes wrong to act in ways that make the outcome best, impersonally

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

evaluated. Therefore, for act consequentialism to accommodate a constraint (p. 192) against breaking promises there must be some occasion on which a single token act of promise-breaking would have the following seemingly incompatible features:

i. Its outcome would be worse overall than the outcome of keeping the promise (this act would be wrong according to act consequentialism).
ii. Its outcome would be better overall, impersonally evaluated, than the outcome of keeping the promise (this act would violate a constraint against promise-breaking).

As this suggests, to explain constraints, act consequentialists must distinguish between different senses of "better overall." Breaking the promise on this occasion must be better overall in an impersonal sense, according to the idea of a moral constraint. But it must also be worse overall, to be wrong according to act consequentialism. So the sense in which it is worse overall must not be the (same) impersonal one.

The way that act consequentialists try to reconcile these claims is by introducing a nonimpersonal, or "agent-relative," way of evaluating outcomes. They can say that what matters (at least some of the time) for an agent's reasons, and the rightness of her actions, is the agent-relative value of the outcomes of her actions, not their impersonal value. For example, they can say that, when it comes to keeping promises, actions of the kind *the agent herself breaking promises* are worse than actions of the kind *others breaking promises*. They can then say that it is worse overall, in this agent-relative sense, for her to break the promise, even though it is better overall, impersonally, for her to do so. By adopting a suitable agent-relative theory of goodness, act consequentialists can try to explain moral constraints (Portmore 2011, chap. 4).[21]

Indirect consequentialists, such as rule consequentialists, can offer a different sort of explanation. Since their theories are not extensionally equivalent to act consequentialism, they anyway imply that it can be wrong to perform some action even when it would have the best outcome overall. That basic feature of the idea of moral constraints fits easily into the structure of indirect consequentialist theories. A more pressing question is whether it is plausible that the best set of rules, or the best conscience or set of motives, corresponds with the moral constraints that we wish to explain.

We can approach this question using the hypothesis that indirect theories employ the idea of pattern-based reasons. Recall that the idea is that an agent S has a pattern-based reason to perform action A in circumstances C if and only if and because S could do A in C and A is S's part, in C, of some favored and eligible pattern of action P. If we were to ask, for example, whether an indirect theory can explain a moral constraint against breaking promises, we could treat this as being the question whether it can explain the eligibility and value of some pattern P in which the agent's part in the relevant circumstances is not to break a promise. This pattern might be one in which no agent ever breaks a promise, or it could be something more complex. If a plausible explanation of the eligibility of some suitable pattern can be given, then an indirect theory could (p. 193) explain the idea that there is a reason not to break promises, even when doing so makes the outcome better.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The project of explaining constraints illustrates some basic theoretical choices facing consequentialist theories of reasons. Act consequentialism has a simple structure, and so any explanation it gives of the complexity of reasons requires a complex theory of the value of outcomes. Rule consequentialism, along with other indirect theories, has a more complex structure, and so it can retain a relatively simple theory of the value of outcomes while seeking to explain the complexity of reasons. Each approach has its own attractions and faces its own challenges.

For example, it is somewhat plausible to say that there is special disvalue in the agent breaking a promise herself, so far as her reasons go, as compared with others breaking promises. So the act consequentialist explanation of a constraint against promise-breaking is somewhat plausible. But the same kind of explanation may be less plausible for other constraints, such as a constraint against torture. Intuitively, the badness of torture is mostly a matter of the suffering caused to the victim, which is bad in an agent-neutral way. An act consequentialist explanation of a constraint against torture would have to claim, instead, that it would be wrong for me to torture someone when doing so would make the outcome better, in an impersonal way, because of the agent-relative badness of my act of torture, or of the relationship with my victim this would instantiate. In contrast, indirect theories need not appeal to agent-relative accounts of value, including the badness of torture, to explain constraints. They can say that the constraint against torture is explained by the agent-neutral badness of torture, which explains why a pattern involving the agent torturing someone is disfavored compared to other eligible patterns. On the other hand, the challenge facing these theories is to develop a plausible account of the eligibility of patterns.

5. How Do Reasons Interact?

A full theory of reasons would tell us several things about them. For example, it would tell us which reasons agents have, and why. It would also tell us how strong these reasons are. But a further question is how reasons interact with each other, and how their interaction relates to the rightness of actions. We touched on these issues in section 1, but it is worth briefly returning to them.

On a simple picture, an action is right if and only if there is no alternative action for which there is stronger reason. If this simple correlation holds, we would know which actions were right if we knew which reasons were strongest. If strength of reasons also *explains* rightness, as is somewhat plausible, we could identify which actions were right, and explain what makes them right, if we knew which reasons were strongest. This promise of explanatory elegance makes the simple picture attractive. But there are alternative possible pictures, with different merits. Earlier we postulated that the connection between rightness and reasons is that an action is right if there is “sufficient reason” to (p. 194) perform it. The simple picture says, in effect, that there is sufficient reason to perform an action just in case there is no alternative action for which there is a stronger reason. This is to assume that sufficiency of reason requires maximum strength of reason.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

This assumption is somewhat attractive, and it may seem to fit with the spirit of consequentialism.

However, nothing in the way we have defined consequentialism here mandates it. Consider briefly the merits of an alternative picture, according to which there is sufficient reason for an action just in case either there is no alternative action for which there is a stronger reason, or the agent has a moral right to perform the action. This alternative implies that it is never wrong to perform an action which you have a moral right to perform.[22] This would mean that, if consequentialists could account for moral rights, they would also be able to account for moral options. That is, they would be able to account for the idea that there is moral discretion in the exercise of moral rights, in the sense that it is not wrong for an agent to perform any of some set of alternative actions which would not have equally good outcomes. Since options (like constraints) are features of "common-sense morality," it is an important question whether consequentialists are able to account for them.

As mentioned briefly in section 1, a further issue to consider is whether some kinds of reasons always or sometimes take priority over other kinds of reasons. For example, we might think that moral reasons sometimes but not always take precedence over other kinds of reason. If moral reasons are sometimes defeated by other reasons, what is right overall would diverge from what is morally right. In these and other ways, consequentialist theories of reasons can try to account for our convictions about the complexity of agents' reasons for action (Portmore 2011, chap. 5; Sobel 2007, 14–17).

6. Conclusion

The topic of consequentialist theories of reasons for action is underexplored when compared to the topic of consequentialist theories of the rightness of actions. Developing theories of reasons more fully would enable consequentialists to account for a greater part of what we care about in ethics. It would also create additional opportunities to introduce complexity into consequentialist ethical theories. Since common criticisms of consequentialism often involve the idea that it is too simple to account for the messiness of ethics, these opportunities are certainly worth exploring.

Developing consequentialist theories of reasons more fully should be of interest to non-consequentialists as well. It is often said that you do not have to be a consequentialist to think that the fact that an action would lead to a better outcome, for example, is (at least sometimes) a reason to perform it. Arguably, the same is true of the fact that an (p. 195) action would be your part in some beneficial or harmful pattern of action. Anyone who believes in either kind of reason has an interest in deepening our understanding of them.[23]

References

Alvarez, M. 2018. "Reasons for Action, Acting for Reasons, and Rationality." *Synthese* 195: 3293–3310.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Bacharach, M. 1999. “Interactive Team Reasoning: A Contribution to the Theory of Co-operation.” *Research in Economics* 53: 117–147.

Bardsley, N., Mehta, J., Starmer, C., and Sugden, R. 2010. “Explaining Focal Points: Cognitive Hierarchy Theory *Versus* Team Reasoning.” *The Economic Journal* 120: 40–79.

Bradley, B. 2018. “Contemporary Consequentialist Theories of Virtue.” In *The Oxford Handbook of Virtue*, edited by N. E. Snow, 398–412. Oxford: Oxford University Press.

Brandt, R. 1992. *Morality, Utilitarianism, and Rights.* Cambridge: Cambridge University Press.

Chappell, R. Y. 2015. “Value Receptacles.” *Noûs* 49, no. 2: 322–332.

Crisp, R. 2006. *Reasons and the Good.* Oxford: Clarendon Press.

Dietz, A. 2016. “What We Together Ought to Do.” *Ethics* 126: 955–982.

Foot, P. 1985. “Utilitarianism and the Virtues.” *Mind* 94, no. 374: 196–209.

Forcehimes, A., and Semrau, L. 2018. “Are There Distinctively Moral Reasons?” *Ethical Theory and Moral Practice* 21: 699–717.

Hooker, B. 2000. *Ideal Code, Real World.* Oxford: Clarendon Press.

Kagan, S. 1989. *The Limits of Morality.* Oxford: Clarendon Press.

Kagan, S. 2000. “Evaluative Focal Points.” In *Morality, Rules, and Consequences: A Critical Reader*, edited by B. Hooker, E. Mason, & D. Miller, 134–55. Edinburgh: Edinburgh University Press.

Lord, E. 2015. “Acting for the Right Reasons, Abilities, and Obligation.” In *Oxford Studies in Metaethics*, Vol. 10, edited by R. Shafer-Landau, 26–52. Oxford: Oxford University Press.

Louise, J. 2004. “Relativity of Value and the Consequentialist Umbrella.” *The Philosophical Quarterly* 54, no. 217: 518–536.

Lyons, D. 1980. “Utility as a Possible Ground of Rights.” *Noûs* 14, no. 1: 17–28.

McElwee, B. 2010. “The Rights and Wrongs of Consequentialism.” *Philosophical Studies* 151: 393–412.

Norcross, A. 1997. “Good and Bad Actions.” *Philosophical Review* 106, no. 1: 1–34.

Parfit, D. 1987. *Reasons and Persons.* Oxford: Clarendon Press.

Parfit, D. 2011. *On What Matters.* Vol. 1. Oxford: Oxford University Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Podgorski, A. 2018. “Wouldn’t It Be Nice? Moral Rules and Distant Worlds.” *Noûs* 52, no. 2: 279–294.

Portmore, D. W. 2011. *Commonsense Consequentialism: Wherein Morality Meets Rationality.* New York: Oxford University Press.

Portmore, D. W. 2019. *Opting for the Best: Oughts and Options.* Oxford: Oxford University Press.

(p. 196) Ridge, M. 2006. “Introducing Variable Rate Rule Utilitarianism.” *The Philosophical Quarterly* 56, no. 223: 242–53.

Roberts, M. A., and Wasserman, D. T., eds. 2009. *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem.* Dordrecht, the Netherlands: Springer.

Ross, W. D. [1930]. 2002. *The Right and the Good*. Edited by P. Stratton-Lake. Oxford: Clarendon Press.

Scanlon, T. M. 1998. *What We Owe to Each Other.* Cambridge, MA: Belknap Press.

Scheffler, S. 1982. *The Rejection of Consequentialism.* Oxford: Clarendon Press.

Scheffler, S. 1992. *Human Morality*. New York: Oxford University Press.

Schroeder, M. 2007. “Teleology, Agent-Relative Value, and ‘Good.’” *Ethics* 117: 265–295.

Smith, H. M. 2010. “Measuring the Consequences of Rules.” *Utilitas* 22, no. 4: 413–433.

Snedegar, J. 2017. *Contrastive Reasons*. Oxford: Oxford University Press.

Sobel, D. 2007. “The Impotence of the Demandingness Objection.” *Philosophers’ Imprint* 7, no. 8: 1–17.

Star, D., ed. 2018. *The Oxford Handbook of Reasons and Normativity.* Oxford: Oxford University Press.

Ullmann-Margalit, E., and Morgenbesser, S. 1977. “Picking and Choosing.” *Social Research* 44, no. 4: 757–785.

Way, J., and Whiting, Daniel. 2017. “Perspectivism and the Argument from Guidance.” *Ethical Theory and Moral Practice* 20: 361–374.

Woodard, C. 2009. “What’s Wrong with Possibilism.” *Analysis* 69, no. 2: 219–226.

Woodard, C. 2013. “The Common Structure of Kantianism and Act Utilitarianism.” *Utilitas* 25, no. 2: 246–265.

Woodard, C. 2019. *Taking Utilitarianism Seriously.* Oxford: Oxford University Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Notes:

([1]) Many who discuss consequentialism take directness of evaluation to be a defining feature. This assumption appears to lie behind the worry that rule consequentialism is incoherent, for example.

([2]) A recent collection of essays on reasons (including normative reasons for action) is edited by Star (2018).

([3]) In Scanlon's influential formulation, a normative reason in general is "a consideration that counts in favor of" something (Scanlon 1998, 17; see also Crisp 2006, 38; and Parfit 2011, 31–33).

([4]) For a recent discussion of some different views about the nature of motivating reasons

see Alvarez (2018).

([5]) For a defense of perspectivism, see Lord (2015). For criticism of one prominent argument for perspectivism, see Way and Whiting (2017). Some of these issues are explored in Woodard (2019, chap. 3).

([6]) Brian McElwee (2010, 397) claims that wrongness is more closely related to blameworthiness than to reasons.

([7]) Standardly, "right" is taken to be ambiguous between "required" and "optional." An action is required if and only if it is wrong not to perform it. An action is optional if and only if it is not wrong to perform it, and not wrong not to perform it.

([8]) There might be sufficient reason to take either can rather than none, of course. The example is drawn from Ullmann-Margalit and Morgenbesser (1977, 761). See also Snedegar (2017).

([9]) For discussion, see Scheffler (1992, chap. 4); Portmore (2011, chap. 2); and Forcehimes and Semrau (2018).

([10]) Not just any theory of goodness can be combined with consequentialism. To avoid circularity, consequentialists cannot appeal to a theory of goodness that makes essential reference to the ethical phenomenon the consequentialist seeks to explain.

([11]) This is to claim that whether a reason exists depends on the features of relevant alternative outcomes. This is compatible with, but does not entail, "contrastivism" about reasons. Contrastivism is the claim that the reason relation contains an argument place for alternatives. See Snedegar (2017, 7–8).

([12]) Should this read "better than the corresponding features of *all* relevant alternative outcomes"? No: if you could leave Tom in misery, or make him happy, or make him very happy, you have some reason to make him merely happy—even though it would be wrong to do so, according to act consequentialism.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([13]) Alternatively, act consequentialists could offer some account of "sufficient reason" that does not refer to goodness overall. If that were possible, then they could formulate their theory without mentioning goodness overall. Individual good features of outcomes would explain reasons, and rightness could be explained directly in terms of the notion of sufficient reason. This would provide one way for act consequentialists to respond to worries about the concept of goodness overall, such as those expressed in Foot (1985).

([14]) Brad Hooker (2000, 88–92) appears to think of them in this way.

([15]) As this suggests, we can think of the reasons posited by act consequentialism as a limiting case of pattern-based reasons, in which the "favored pattern" is identical to the agent's act. See Bacharach (1999, 118). We can also think of the actualism vs. possibilism debate in deontic logic in terms of pattern-based reasons: see Woodard (2009).

([16]) Another example is the rule requiring beneficence that Hooker postulates (2000, 98n7). This is a general reason to benefit others, though it is subordinate to other rules in the ideal code.

([17]) This is to claim that whether a pattern-based reason exists depends on the features of the outcomes of relevant alternative patterns. This is parallel to act consequentialism's claim that whether a reason exists depends on the features of the outcomes of relevant alternative actions.

([18]) Motive consequentialism also faces issues of embedding. For relevant discussion, see Kagan (2000) and Bradley (2018).

([19]) Scheffler calls constraints "agent-centred restrictions."

([20]) This example is drawn from Lyons (1980, 17–28).

([21]) As is well-known, the theory of goodness would also have to be time-relative.

([22]) More complex accounts of sufficient reason are also possible. For example, it is possible that it is usually right to perform an action for which you have a moral right, unless there is a *much* stronger reason to do something else.

([23]) I am very grateful to Douglas Portmore and Penelope Mackie for very helpful comments on a draft of this chapter.

Christopher Woodard

Christopher Woodard is Professor of Moral and Political Philosophy at the University of Nottingham, UK, and President of the British Society for Ethical Theory. His research focuses on consequentialism (especially collective forms of consequentialism), well-being, and normative reasons for action. He has also written on other topics in moral and political philosophy, including egalitarianism, meaning in life, and the actualism-possibilism debate. He is the author of two books: Reasons, Patterns, and Co-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

operation (Routledge, 2008) and Taking Utilitarianism Seriously (Oxford University Press, 2019).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Consequentialism and Nonhuman Animals

Tyler M. John and Jeff Sebo

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.32

Abstract and Keywords

Consequentialism is thought to be in significant conflict with animal rights theory because it does not regard activities such as confinement, killing, and exploitation as in principle morally wrong. Proponents of the "Logic of the Larder" argue that consequentialism results in an implausibly pro-exploitation stance, permitting us to eat farmed animals with positive well-being to ensure future such animals exist. Proponents of the "Logic of the Logger" argue that consequentialism results in an implausibly anti-conservationist stance, permitting us to exterminate wild animals with negative well-being to ensure future such animals do not exist. We argue that this conflict is overstated. Once we have properly accounted for indirect effects, such as the role that our policies play in shaping moral attitudes and behavior and the importance of accepting policies that are robust against deviation, we can see that consequentialism may converge with animal rights theory significantly, even if not entirely.

Keywords: animal rights, conservation, farmed animal welfare, wild animal welfare, two-level utilitarianism, effective altruism, meat paradox, population ethics

1. Introduction

CONSEQUENTIALIST moral theories and nonhuman animals share a long and complicated history.[1] On one hand, some of the earliest Western philosophers to take seriously the moral status of nonhuman animals were the British utilitarians Jeremy Bentham, J. S. Mill, and Henry Sidgwick. Moreover, contemporary utilitarian Peter Singer is often credited with having started the modern-day animal rights movement with the publication of *Animal Liberation*. Consequentialist principles motivate many animal advocates in general, and they are at the foundation of the effective animal advocacy movement in particular.[2]

On the other hand, many philosophers and advocates question whether consequentialism adequately respects the moral status of nonhuman animals, in much the same way that they question whether consequentialism adequately respects the moral status of humans. Familiar critiques of consequentialism emerge with new life in the context of assessing

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

the moral status of nonhuman animals, such as the critique that consequentialism regards individuals as fungible receptacles of value, which is to be promoted regardless of the means of its promotion.

In this chapter we will focus on two related issues that arise for consequentialists regarding nonhuman animals, one regarding domesticated animals and the other regarding (p. 565) wild animals. Regarding domesticated animals, some philosophers believe that consequentialism results in an implausibly pro-exploitation stance, according to which, if farmed animals have positive well-being, then we are morally permitted if not required to increase the number of farmed animals in the world, all else equal. Regarding wild animals, some philosophers believe that consequentialism results in an implausibly anti-conservationist stance, according to which, if wild animals have negative well-being, then we are morally permitted if not required to decrease the number of wild animals in the world, all else equal.

This chapter assesses whether standard forms of consequentialism have these results. Our approach echoes arguments from numerous consequentialist writers before us, such as Henry Sidgwick and R. M. Hare. We should make a distinction between *criteria of rightness*, which determine which actions are right in theory, and *decision procedures*, which we use to decide which actions to perform in practice.[3] When we do, we find that consequentialism as a criterion of rightness recommends a partly consequentialist, partly nonconsequentialist decision procedure for most people in most situations. In our view, this partly consequentialist, partly nonconsequentialist decision procedure conflicts with pro-exploitation and anti-conservation stances. Thus, we will argue, the consequentialist case for abolition of animal agriculture and conservation of wild animal habitats is stronger than many philosophers appreciate.

Before we begin, we should make some caveats about the scope of our discussion. First, there are many normative questions about which consequentialists disagree, some which bear on the topics that we discuss. These questions include: Should we accept hedonism, desire satisfactionism, or something else as our theory of the good? Should we accept act consequentialism, rule consequentialism, or something else as our theory of the right? And so on. We will not be able to discuss all these issues here. Instead, we will focus on classical utilitarianism (i.e., actualist, hedonist, maximizing,[4] totalist, act consequentialism), and we will note issues about which different consequentialist theories have different implications.

Second, there are many empirical questions about which consequentialists disagree, some of which bear on the topics that we discuss as well. For example, do farmed animals and wild animals in fact have positive or negative well-being? Does our individual behavior make a difference regarding how many farmed animals or wild animals are in the world? And so on. Once again, we will not be able to discuss all these issues here. Instead, we will stipulate answers to these questions for the sake of discussion where necessary, and we will allow these questions to remain open where possible. In all cases, we

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

will do our best to note these questions where they arise and to explain why we approach them in the way that we are.

(p. 566) Third, and relatedly, we will not provide a conclusive answer to the questions we are considering. How many domesticated and wild animals there should be, and what we should do in order to realize these population levels, are extraordinarily complicated questions that require comprehensive normative and empirical analysis to answer. Instead, we will do the following. First, we will situate nonhuman animals in consequentialist theory. Second, we will summarize and evaluate arguments that philosophers have made regarding consequentialism, farmed animals, and wild animals. Third, we will introduce a set of considerations that we take to provide strong, and possibly decisive, support for abolitionist and conservationist stances from a consequentialist perspective.

2. Background

2.1. Situating Animals in Consequentialist Theory

We take consequentialism to be a family of moral theories according to which the rightness of actions is entirely a function of their consequences. Philosophers disagree widely about the scope of this family.[5] As such, and for the sake of simplicity and specificity, we focus on paradigmatic forms of consequentialism which are impartially benevolent and which reject the act/omission distinction and other standard deontological distinctions.

Understood in this way, consequentialism has historically been a more species-egalitarian family of moral theories than its competitors. This is partly due to the influence of classical utilitarians, who appreciated that a principled, impartially benevolent, welfarist moral theory implies that all sentient beings have equal moral standing. As Bentham famously stated, “The question is not, Can they reason?, nor Can they talk? but, Can they suffer?”[6] By contrast, nonconsequentialist theorists such as Kantians and contractualists have for the most part only recently begun to accept that nonhumans can have moral standing at all.[7] Our view is that the historically consequentialist view is correct. We therefore assume throughout that *all animals are equal*, in the sense that all animals’ interests merit equal moral consideration.

Much of the modern-day project of determining how to maximize impartial good is taken up by the effective altruism community. Effective altruism is, broadly, the project of using evidence and reason to determine how to improve lives as much as possible, and then acting accordingly.[8] While effective altruism is compatible with other moral theories, many people see it as characteristically consequentialist. This is partly because consequentialists such as Toby Ord and Peter Singer developed the idea of effective altruism, and partly because the idea of effective altruism focuses centrally on maximizing good outcomes.[9]

(p. 567) Effective altruists assess the priority of different focus areas using three heuristics: importance, tractability, and neglectedness.[10] A problem is more important to the extent that solving it would make a positive difference to the world. A problem is more

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

tractable to the extent that it is easy to solve. A problem is more neglected to the extent that few people are working on solving it. While there are important limitations to this framework,[11] when properly applied it serves as a useful guide to identifying the problems that are, in consequentialist terms, the most important problems to address.

Using the importance, tractability, neglectedness framework, effective altruists have identified three major areas as among the highest-priority cause areas for altruistic intervention: animal welfare, global health and development, and existential risk reduction. Moreover, within animal welfare, effective altruists think that farmed animal welfare and wild animal welfare are the highest-priority issues.

Consider farmed animal welfare first. This issue is highly important due to its immense scale: we harm 100+ billion domesticated animals and hundreds of billions of wild animals per year in our global food system.[12] This issue is also highly neglected: people devote much less time, energy, and money to farmed animal welfare than to other issues, such as companion animal welfare. Finally, this issue is also highly tractable: people are currently pursuing a variety of promising approaches involving social, institutional, political, and technological change.

Now consider wild animal welfare. This issue is even larger in scale than farmed animal welfare: anywhere between 10^{13}–10^{16} vertebrates and 10^{18}–10^{22} invertebrates live in the wild at any given time, many with low levels of well-being.[13] This issue is also even more neglected than farmed animal welfare: hardly anyone is working on it at all. However, wild animal welfare is not nearly as tractable as farmed animal welfare, since we currently lack the political will to promote wild animal welfare as well as knowledge about what we can do to efficiently improve the lives of wild animals.[14]

While effective altruists agree that farmed animal welfare is more tractable than wild animal welfare, there are many uncertainties with respect to both issues. With respect to farmed animal welfare, we need to know whether to aim to abolish or regulate animal agriculture, as well as how to pursue these ends. With respect to wild animal welfare, we need to know whether to aim to increase, decrease, or maintain wild animal populations, as well as how to pursue these ends. In both cases, we need to strike a balance between a willingness to be humble in the face of difficult questions and a willingness to be proactive with respect to urgent issues.

2.2. Sophisticated Consequentialism

We believe that, as consequentialists think about how to answer these questions, it is important to appreciate the distinction between (a) *criteria of rightness*, that is, the principles that determine which actions are right in theory, and (b) *decision procedures*, that (p. 568) is, the principles that agents use to decide which actions to perform in practice.[15] This distinction is important because, as many consequentialists have observed, it might not always be the case that consulting a particular principle, such as the principle of utility, is the best way to comply with that principle.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

There are many reasons why the decision procedures we ought to use might be different from our criteria of rightness. One reason concerns complexity. Insofar as we lack the time, energy, and information necessary to apply complex principles, we should apply simpler principles instead. Another concerns biases and heuristics. Insofar as complex principles create more space for bias to operate, we should apply simpler principles instead. Another concerns moral psychology. Insofar as our behavior depends on factors other than explicit moral reasoning, we should attend to these factors as well. And so on.

With that in mind, our view, stated roughly and generally, is that consequentialist theorists who have defended so-called indirect consequentialism, sophisticated consequentialism, or two-level consequentialism are correct.[16] Classical utilitarianism is correct as criterion of rightness: we morally ought to perform the acts which maximize net pleasure for all sentient beings from now until the end of time. However, for most people in most situations, a partly consequentialist and partly nonconsequentialist framework is the optimal decision procedure. According to this kind of decision procedure, we should aim to maximize expected utility, but only where this is compatible with respecting rights, developing and maintaining relationships of care, and developing and maintaining virtuous character traits. While different decision procedures may be optimal for different people in different contexts, decision procedures of this kind generally strike a good balance between (a) preserving the benefits of consequentialist thinking and (b) limiting the risks of consequentialist thinking.

With that said, we should qualify this claim in two ways. First, we are open to the possibility that we are wrong. After all, these are difficult questions, and biases and heuristics can affect our application of any decision procedure. For example, once we accept that we should accept a partly nonconsequentialist decision procedure, it might be tempting to simply select whatever decision procedure tells us what we want to hear, and then rationalize our choice on specious consequentialist grounds. We will not be able to fully address this concern here, but we will note where it might be arising, and we will approach our own analysis with a degree of skepticism accordingly.

Second, we suspect that, even if we are right, there can be exceptional cases where a fully consequentialist decision procedure which suspends nonconsequentialist constraints is best. For example, it might be that an optimal decision procedure would allow you to decide to kill someone if doing so is the only way to save 1,000,000 people, even though you should ordinarily regard killing someone as prohibited on nonconsequentialist grounds. In this case, you would not be denying the indirect value of nonconsequentialist **(p. 569)** considerations. You would instead simply be accepting that the nature of this case makes it clear that a fully consequentialist decision procedure is ideal. However, we think that cases of legitimate suspension of nonconsequentialist constraints as weighty as rights are rare, and they may not occur at all for many people.

We think that this kind of “sophisticated consequentialism” has interesting implications for a wide range of issues in animal ethics. For example, we think that it implies that we should support the development of a broad, pluralistic animal advocacy movement that

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

involves many different, and seemingly conflicting, approaches.[17] In what follows, we will focus on implications regarding how many farmed and wild animals there should be in the world and what we should be doing to promote these population levels. Without attempting to fully answer these questions here, we will argue that there is a stronger consequentialist case for abolition of animal agriculture and conservation of wild animal habitats than many philosophers assume.

3. Farmed Animals and the Logic of the Larder

3.1. Background

The standard argument that consequentialists should aim to reduce farmed animal populations, all else equal, relies on the assumption that farmed animals have net negative well-being. At least in countries with developed, industrialized economies, which will be our focus, there are good reasons for embracing this assumption. For brevity, consider the fates of farmed chickens, who make up over 99 percent of the population of farmed land animals in the United States. Approximately 99.9 percent of chickens farmed for meat and 98.2 percent of chickens farmed for eggs live in concentrated animal feeding operations (CAFOs).[18] Lori Gruen writes on the lives of such chickens:[19]

> Most of these hens are kept in small wire cages, called “battery cages,” with between three and eight other hens. The battery cages are stacked on top of each other indoors in sheds that can contain upward of 100,000 hens. The battery cage is so small that the hens are unable to stretch their wings or turn around. Because of the stress, boredom, fear, and close quarters, hens will peck at each other, so most are routinely debeaked, a process that involves a hot blade cutting off the tip of the beak through a thick layer of highly sensitive tissue. Debeaking causes lasting pain and impairs the hen’s ability to eat, drink, wipe her beak, and preen normally.

(p. 570) Many other impacts reduce chicken well-being as well—the pain and stress of laying each of 300 eggs per year, an inability to stand due to rapid growth leading to chronic leg pain and constant sores from sitting in their own excrement, and more—and even setting these aside it is clear that animals raised on such CAFOs have profoundly negative well-being.

However, even if the vast majority of farmed animals have negative well-being, there may be some farmed animals who presently exist (such as some grass-fed “beef” cattle) or who might exist in the future (such as genetically engineered, pain-free chickens) who have neutral or positive well-being. Dwelling on such cases has led some to defend the so-called Logic of the Larder (hereafter LARDER):[20]

> [Where farmed animals have positive well-being,] the consequence to others of buying that meat in the grocery store, rather than asparagus, is good; you create farm animals whose lives are worth living. … So if you, like me, think your actions

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> are more moral when you do more good for others, you should agree with me that [this] meat is moral, and veggies are immoral.

The idea here is that, if consequentialism is true, and if some farmed animals have positive well-being, then there is a pro tanto moral reason to promote a world that includes these farmed animals instead of a world that excludes them.[21] This might mean that we have pro tanto moral reasons to eat animal products that come from such a farm and to support the existence of such a farm in other ways. Many defenders of LARDER further suppose that these moral reasons are ultimately undefeated, such that, all things considered, consequentialists ought to eat some "humanely raised" meat.

The weakest version of LARDER, on which eating some farmed animals is *permissible* because it is *not bad* to cause farmed animals with positive well-being to exist, makes weak assumptions about population axiology. In particular, it assumes a weak version of the Mere Addition Principle:[22] that adding animals with positive well-being to our actual world does not make the world worse, holding everything else fixed. It is not committed to rejecting the Asymmetry Intuition, or critical level or averageist axiologies, though each of these axiologies will change the conditions under which adding animals with positive well-being to the world would not worsen that world. The strongest version of LARDER, supported by classical utilitarianism, implies that eating some (p. 571) farmed animals is *required* because it is *good* to cause farmed animals with positive well-being to exist.

While many people writing on LARDER have focused on its implications for the ethics of eating animals, it is clear that the argument has broader implications for our relationships with nonhuman animals. If consequentialism requires agents to take actions which increase the number of farmed animals with positive well-being, all else equal, then it might require us to support animal agriculture in other ways, too, for example by aiming to regulate rather than abolish animal agriculture as an industry. Whereas animal rights theory regards animal farming as anathema, consequentialism on this interpretation might regard it as welcome.

Some philosophers thus reply to LARDER by rejecting consequentialism. They claim that supporting animal agriculture is wrong whether or not farmed animals have positive well-being, on the grounds that animal agriculture treats animals merely as means, cultivates vicious attitudes toward animals, or places us in oppressive relationships with animals.[23]

Other philosophers reply to LARDER by rejecting the idea that consequentialism supports increasing farmed animal populations. For example, Matheny and Chan argue that supporting animal agriculture is unlikely to maximize value all things considered, since other uses of our time, energy, and money will have better net consequences.[24]

Other philosophers reply to LARDER by accepting the idea that consequentialism supports increasing farmed animal populations. If engaging in or supporting animal agricul-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ture is a net benefit for farmed animals, then we are indeed morally permitted, if not morally required, to engage in or support animal agriculture, all else equal.

We are sympathetic with all of these replies. First, we agree with nonconsequentialist critics of LARDER that we should treat animals as ends, cultivate virtuous character traits toward animals, and cultivate relationships of care with animals. However, we think that we should do these things from within a consequentialist framework—because doing these things maximizes net pleasure in the world—rather than as an alternative to a consequentialist framework.

Second, we agree with consequentialist critics of LARDER that animal agriculture is unlikely to be a net benefit for farmed animals in practice. However, we think that there is a deeper reason for consequentialists to reject LARDER, which is that even treating LARDER as an open question is likely to be a net harm for nonhuman animals and other sentient beings in most cases in practice, for precisely the reasons that nonconsequentialists are discussing.[25]

Third, we agree with consequentialist proponents of LARDER that, if animal agriculture is a net benefit for farmed animals and other sentient beings, then we are morally permitted, if not morally required, to support animal agriculture, all else equal, in theory. (p. 572) However, we also think that we are not morally permitted to support animal agriculture in most cases in practice, again for the reasons that nonconsequentialists are discussing.

Our aim in what follows, then, is to argue that a consequentialist criterion of rightness requires us to accept a partly nonconsequentialist decision procedure, and that this decision procedure prohibits eating animals, as well as maintaining and supporting systems that confine, kill, and exploit animals as a matter of principle (with certain caveats that we explain). This is centrally because supporting animal agriculture negatively shapes our individual beliefs, values, and practices, and because having a system of animal farming at all negatively shapes our collective beliefs, values, and practices. In both cases, the result is that we tend to have attitudes that devalue animals and practices that harm them.

3.2. The Individual Effects of Animal Exploitation

We begin with the individual effects of animal exploitation. We here follow the literature in focusing on the psychological effects of eating meat, though we will consider later whether and to what degree these effects apply to other activities that involve exploitation, too.

Our argument has two parts. First, theoretical and empirical moral psychology support the idea, originally found in ecofeminist thought, that eating animals leads humans to view animals as having diminished mental life and moral status. When we condone animal agriculture, in word, thought, or deed, we condition ourselves to devalue and, as a result, harm other animals. Second, theoretical and empirical motivational psychology sup-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ports the idea that so-called conscientious omnivores typically fail to be as conscientious as they would like to think. That is, when we adopt a policy of eating happy animals, we will likely end up eating unhappy animals as well. Thus, we will argue, consequentialists should adopt a policy of not eating animals at all (with certain caveats that we discuss).[26]

Part one of our argument—that eating animal products conditions us to see animals as objects rather than subjects—has precedent among consequentialists and nonconsequentialists alike. For example, Peter Singer argues:[27]

> [Practically], it would be better to reject altogether the killing of animals for food, unless one must do so to survive. Killing animals for food makes us think of them as objects that we can use as we please. … To foster the right attitudes of consideration (p. 573) for animals … it may be best to make it a simple principle to avoid killing them for food.

Similarly, Cora Diamond points out that humans reject emphatically the practice of eating our own dead, not because we think that we have a moral duty not to engage in this practice, but rather because we have relationships with and attitudes toward humans in light of which it simply makes no sense to eat them. To eat a human body is to commit a kind of category error. Committing this error expresses a kind of disregard for the miscategorized subject, by placing them in the category of the edible rather than in the category of the personal.[28]

Building on Diamond's line of argument, Lori Gruen has argued that what is wrong with eating animals is that:[29]

> [I]n turning other animals from living subjects with lives of their own into commodities or consumable objects we have erased their subjectivity and reduced them to things … [This] forecloses another way of seeing animals, as beings with whom we can empathize and learn to understand and respond to differences.

Finally (though there are other examples too), Carol Adams argues that:[30]

> [M]eat-eating offers the grounds for subjugating animals: if we can kill, butcher, and consume them—in other words, completely annihilate them—we may as well experiment upon them, trap and hunt them, exploit them, and raise them in environments that imprison them, such as factory and fur-bearing animal farms.

Recent psychological research on the so-called meat paradox empirically confirms these claims. For example, in a series of five studies, Brock Bastian and colleagues have demonstrated a link between seeing animals as food, on one hand, and seeing animals as having diminished mental lives and moral value, on the other hand. We will here describe three.

In a first study, participants were asked to rate the degree to which each of a diverse group of thirty-two animals possessed ten mental capacities, and then were asked how likely they would be to eat the animal and how wrong they believe eating that animal is. Perceived edibility was negatively associated with mind possession ($r = –.42$, $p < .001$),

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

which was in turn associated with how the perceived wrongness of eating the animal (r = .80, p < .001).[31]

In a second study, participants were asked to eat dried beef or dried nuts and then judge a cow's cognitive abilities and desert of moral treatment on two seven-point scales. Participants in the beef condition (M = 5.57) viewed the cow as significantly less deserving of moral concern than those in the control condition (M = 6.08).[32]

(p. 574) In a third study, participants were informed about Papua New Guinea's tree kangaroo and informed variably that tree kangaroos have a steady population, that they are killed by storms, that they are killed for food, or that they are foraged for food. Bastian and colleagues found that categorizing tree kangaroos as food and no other features of these cases led participants to attribute less capacity for suffering and less moral concern.[33]

Additionally, a sequence of five studies from Jonas Kunst and Sigrid Hohle demonstrates that processing meat, beheading a whole roasted pig, watching a meat advertisement without a live animal versus one with a live animal, describing meat production as "harvesting" versus "killing" or "slaughtering," and describing meat as "beef/pork" rather than "cow/pig" all decreased empathy for the animal in question and, in several cases, significantly increased willingness to eat meat rather than an alternative vegetarian dish.[34]

Psychologists involved in these and several other studies[35] believe that these phenomena occur because people recognize an incongruity between eating animals and seeing them as beings with mental life and moral status, so they are motivated to resolve this cognitive dissonance by lowering their estimation of animal sentience and moral status. Since these affective attitudes influence the decisions we make—from our consumer behavior to our voting behavior, political advocacy, career choice, philanthropic activity, conversations we have with others, and more— eating meat and embracing the idea of animals as food negatively influences our individual and social treatment of nonhuman animals.

Part two of our argument—that eating animal products in exceptional cases makes us likely to eat animal products in ordinary cases—has precedent as well. Recall that a central reason why Hare and other consequentialists support simpler decision procedures is that more complex decision procedures have more adjustable parameters that allow for false rationalization.

Following this line of reasoning, we can predict that a policy of not eating animal products at all will generally be better than a policy of eating animal products only in narrowly circumscribed contexts. Self-identified "conscientious omnivores" who claim to eat animal products only in circumstances where farmed animals have positive well-being are likely to eat animal products in circumstances where farmed animals have negative well-being as well. In particular, they are likely to rationalize eating animal products not only on the grounds that animals experience diminished pain or have diminished moral status, but also on other grounds, such as that they are at a family dinner, that a particular

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

restaurant probably has ethical practices, or even that a particular item on the menu looks appealing.

Here, again, psychological research supports armchair theory. A 2015 study revealed that "conscientious omnivores" were less likely than vegetarians to perceive their diet as something that they needed to follow. They reported violating their diet more, feeling less guilty when doing so, feeling less disgusted by factory-farmed meat, and believing (p. 575) less in animal rights, among other findings.[36] Moreover, diet had a statistically significant effect on all measures independent of whether the diet was motivated by health or ethical reasons. Whether one is a vegetarian or a conscientious omnivore appears to change one's psychological relationship to meat and to meat-eating, with implications for how consistently one applies one's policy. Note also that these self-reports are unlikely to capture cases in which individuals see themselves as complying with their policy when they are in fact violating it, or cases in which individuals see themselves as violating their policy but would rather not admit that.

We are now in a position to see that, even if an individual might be morally permitted to be a "conscientious omnivore" rather than a vegetarian in principle (i.e., in cases that idealize away facts about human psychology), most individuals have strong (in our view decisive) reason not to be "conscientious omnivores" rather than vegetarians in practice (i.e., in cases that do not idealize away facts about human psychology). Because of the indirect effects of conforming to a policy of eating animals sometimes, a policy of not eating animals at all will do more good overall. Thus, consequentialists have strong (in our view decisive) reason to adopt a policy of not eating animals at all, except perhaps in highly exceptional cases where doing so clearly does more good than harm. More generally, we have strong (in our view decisive) reason to adopt a policy of supporting beliefs, values, and practices that treat animals as subjects rather than as objects, and that cultivate relationships of care rather than exploitation with them.

3.3. The Social Effects of Animal Exploitation

We now consider the social effects of animal exploitation. (Here we focus on the social effects of systems of animal exploitation themselves, though we believe that individual support for these systems can have social effects, too.)[37] Once again, our argument has a dual structure, taking the impact of animal farming on human attitudes as one premise and our skepticism about the possibility of restricting ourselves to net positive versions of this practice as another.

Our central contention is that, because animal agriculture is necessarily a system of institutionalized violence against nonhuman animals, the existence of any such system will tend to socially perpetuate a speciesist ideological orientation toward nonhuman animals, diminishing the moral status that society predicates to them. This will, in turn, lead to both systematic violations of compliance with the standards of farming which LARDER requires and to other harmful actions regarding nonhuman animals and other sentient beings.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Animal farming serves as the grounds of its own ideological justification. The very fact that animal farming exists makes us more likely to see it as acceptable, in part by (p. 576) providing us with evidence that other people see it as acceptable. Moreover, the idea that humans can treat nonhumans as we do in animal farming provides an inferential justification for all kinds of other practices and attitudes, including complacency with other systems of nonhuman exploitation and with wild animal suffering. Finally, the production of agricultural imagery—which typically obscures rather than illuminates the realities of animal agriculture because it is funded by industry—establishes animal agriculture as a legitimate and permanent institution.[38]

The idea that a harmful or oppressive system can serve as its own ideological justification is not new. Many people have made this point before, not only in the context of animal rights advocacy but also in the context of human rights advocacy. For example, in her work on prison abolitionism, Angela Davis argues that images of the prison system foster complacency with incarceration. In particular, Davis argues that media productions, especially in Hollywood, make the prison one of the "most important features of our image environment."[39]

> This has caused us to take the existence of prisons for granted. The prison has become a key ingredient of our common sense. It is there, all around us. We do not question whether it should exist. It has become so much a part of our lives that it requires a great feat of the imagination to envision life beyond the prison.

Despite our constant consumption of prisons, the "realities of imprisonment are hidden from almost all who have not had the misfortune of doing time."[40] Cultural images of prisons obscure rather than illuminate the realities of the prison system, all while impressing upon us the necessity, naturalness, and permanence of an expansive system of incarceration. Meanwhile, the prison system functions to racialize punishment, associating Blackness with criminality and with punishment.[41]

Many social and legal theorists believe that the law is similar, in that a central mechanism through which the law yields conformity is by shaping perceived group norms and attitudes, thereby anchoring human moral attitudes and behavior.[42] The law performs this function both directly and indirectly. It performs this function directly when members of a society can infer from changing laws that a certain number of people must support the proposed norm. It performs this function indirectly when members of society view other members following the law and infer that others must endorse the norm which the law enforces.[43]

The upshot is that the system of animal agriculture and the current legal status of animal agriculture work together to socially legitimize this system. They both shape perceived group norms, anchoring our moral attitudes and behaviors. Members of a society can infer from the fact that the system of animal agriculture is legal in that society that (p. 577) most people in that society support confining, killing, and eating animals (and are right to

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

do so). If so, then a legal system of animal agriculture works in multiple ways to justify its own existence, as well as to inferiorize nonhuman animals.

The importance of these effects should not be understated. As some effective altruists argue, some of the very most important interventions that we can perform to improve the total value of the world are aimed at "moral circle expansion." To aim for moral circle expansion is to aim for a wider range of sentient beings to receive moral consideration over time. The idea here is that the values of future generations will make a vast difference to the value of the future—for example, they could change whether these people will support or resist protections for domesticated animals, wild animals, or even digital beings. Moreover, because the number of future nonhuman sentient beings is extremely large in expectation, any difference we can make to the moral behavior of future generations regarding nonhuman sentient beings is astronomical in expected value.[44] Thus, if institutionalized animal agriculture is an obstacle in the way of moral circle expansion, removing this obstacle should be a central moral priority for consequentialists.

Next, notice that a society that maintains a system of animal agriculture in the narrow contexts in which this system is a net benefit for farmed animals will doubtfully be able to contain its farming practices to these contexts. In countries with developed, industrialized economies, animal agriculture manages to produce animal products at scale only by producing them at very low cost to industry. This in turn requires industry to adopt very minimal space requirements, veterinary care, and regulation and oversight, while using genetically modified species whose rapid growth, reproductive efficiency, and hormonal excesses leave them chronically ill and in pain. A system of animal agriculture that provides farmed animals with positive well-being would require drastic revisions to all of these features of animal agriculture, each significantly raising the economic costs of production. While we cannot here build a quantitative model, suffice it to say that we are highly skeptical of the possibility of building a system of animal farming that both benefits farmed animals and feeds anyone beyond the very wealthiest humans.

These concerns might not fully apply to subsistence animal farming with dramatically lower stocking density in countries without developed, industrialized economies. But while this system of animal farming might be able to maintain animal welfare standards conducive to the LARDER over the short term, capitalist selective pressures may eventually favor the development of industrial systems of animal farming to which our concerns will apply fully. Thus, perhaps barring rare cases where animal products are nutritionally mandated, it is plausible that consequentialists should endorse a policy of not farming animals anywhere. With that said, our focus in this chapter is on animal farming in the context of developed, industrialized economies, and so we will not try to argue for this more general policy here.

The upshot of these discussions is that consequentialists have strong reason to reject LARDER at the level of decision procedure. In particular, we should accept principles (p. 578) which forbid increasing and require decreasing the population of farmed animals, and which forbid eating animals and otherwise supporting the idea that animals are food

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

in all but the most exceptional cases. To be clear about the structure of our argument, what these considerations do is raise the moral costs of meat-eating and animal farming. We think that these costs are sufficiently high that the benefits of positive well-being for some farmed animals do not outweigh the costs in all but the most fanciful cases. Once we combine the indirect considerations that we have discussed here with the direct considerations that we discussed earlier (about the expected animal welfare, public health, and environmental impacts of animal agriculture in the real world), the case for abolition of animal agriculture becomes even stronger.

As with any decision procedure, this partly consequentialist, partly nonconsequentialist decision procedure is likely to produce at least some blameless wrongdoing. While eating animals and performatively condoning animal farming will ordinarily be harmful, they might sometimes be beneficial. Moreover, while we might sometimes clearly see when we are in an exceptional case where this is beneficial, we will not always clearly see that. But this is fine. Since no one short of an archangel has the psychological capacity to act optimifically in every choice situation, the best we can do is identify the governing policies that minimize expected wrongdoing over the long run. Our view is that for most people in most situations, this partly consequentialist, partly nonconsequentialist decision procedure does exactly that.

We should note three caveats about our argument here. First, we are not sure to what degree the social and psychological impacts of meat production and consumption extend to other forms of animal use, including the use of animals for eggs, dairy, clothing, research, entertainment, and companionship. We predict that these social and psychological impacts will be strongest in the case of meat production and consumption, but that they will at least be present in the context of other forms of harmful or oppressive use. At the limit, there will be instances of use such as the consumption of plastics made from animal byproducts that are so psychologically divorced from animal use that they may have no individual psychological impacts at all. But this is an empirical hypothesis that requires empirical investigation.

Second, as with any empirical psychological findings, we are not sure to what degree there may be variation in the attitudes toward farmed animals and other sentient beings that people form as a result of consuming animal products and living in a society that uses animals for food. Thus, we are not sure to what degree there is variation in the decision procedures that will help people to maximize net pleasure in the world, given these psychological impacts. The psychological effects that we have discussed in this section appear to be robust, but we should not expect this to be a human psychological universal. Note that since we cannot typically assess our own levels of bias introspectively, we should all assume that we are likely to be subject to the biases described.

Third, we are not sure to what degree there might be exceptional cases where meat production and consumption is morally permissible or required at the decision procedure level. We can at least imagine cases where producing or consuming meat would **(p. 579)** clearly be optimal, such that we should suspend animal rights that we normally regard as

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

absolute. But note that such a case would have to be truly exceptional; that is, it would have to be the kind of case that might warrant suspending human rights as well. Other than cases where people need to produce or consume meat to survive (which are not as common as “conscientious omnivores” think, though they do occur), we expect that such cases will be rare, though we cannot say for sure.

Many people criticize animal advocates for focusing too much on consumer action and not enough on other kinds of political action. We agree with this criticism, which is part of why we recommend advocacy that aims not only at individual consumer change but also at social, political, economic, and technological change. However, we also think that individual consumer change is more important than some critics realize. When we distance ourselves from systems of violence, we are able to see these systems for what they are and to find the motivation to resist them in other ways.

4. Wild Animals and the Logic of the Logger

4.1. Background

The idea that consequentialists should aim to conserve wild animal populations, all else being equal, relies on the assumption that wild animals generally have positive well-being. And it makes sense that people would make this assumption. After all, wild animals do experience positive well-being in their lives. They enjoy food, sex, play, relationships, and a range of comforting solitary and interpersonal experiences.

However, some consequentialists believe that wild animals have negative well-being. Granted, they might have ample opportunity for positive experience. But they also face ample risk of negative experience, resulting from hunger, thirst, illness, injury, predation, and more. Moreover, most wild animals are small animals who are members of “r-selected” species. Such animals achieve population equilibrium by giving birth to very many offspring with extremely high mortality rates. Oscar Horta offers the example of Atlantic Cods, who maintain population equilibrium by spawning around two million eggs per year, only one of which, on average, will reach adulthood. Thus, the vast majority of wild animals who exist, assuming they are sentient, have very short, painful lives that consist mainly of dying.

Such observations have led many commentators to note that if most wild animals have negative well-being, then the world could be improved simply by ending the lives of these animals and destroying their habitats, an argument which we have titled “The Logic of the Logger” (LOGGER). For example, effective altruist blogger Brian Tomasik argues that “[g]iven that most wild animals that are born have net-negative experiences, (p. 580) loss of wildlife habitat should in general be encouraged rather than opposed.”[45] Whereas people like Yew-Kwang Ng encourage “extreme caution before we do anything that may disturb the biosphere,”[46] Tomasik argues that such caution is unwarranted and encourages us to adopt a strong “anti-conservationist” stance.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

The idea here is that if consequentialism is true, and if wild animals have negative well-being, then there is a pro tanto moral reason to promote a world that excludes these wild animals instead of a world that includes them. This might mean that we have pro tanto moral reasons to engage in hunting, fishing, and as Tomasik argues, activities aimed at "decreasing plant growth and entirely eliminating wilderness." Some defenders of the Logic of the Logger further suppose that these moral reasons are ultimately undefeated, such that, all things considered, consequentialists ought to engage in such anti-conservationist activities.

As with LARDER, the weakest version of LOGGER makes conservative assumptions about population axiology. It assumes only that it is *not bad* for there to be fewer wild animals with negative well-being. While there are population axiologies that sometimes deny this, such as averageism and some impartial forms of egalitarianism, the claim that it is not bad for there to be fewer sentient beings with negative well-being is a highly plausible desideratum for population axiology. The strongest version of LOGGER, supported by classical utilitarianism, implies that destroying animals and ecosystems is *required* because it is *bad* for wild animals with negative well-being to exist.

Because LOGGER is a very new argument, discussed mostly on internet blogs and in op-eds, few philosophers have commented on the issue. Those who have commented have made similar responses to LOGGER as to LARDER.[47] In particular, they have replied by rejecting consequentialism, by rejecting the idea that consequentialism supports reducing wild animal populations, and by accepting the idea that consequentialism supports this. Especially important have been arguments that (a) wild animals do not clearly experience net negative well-being,[48] and (b) the possibility of unpredictable trophic cascades makes it difficult if not impossible to identify habitat destruction methods that will do more good than harm overall.[49]

As with LARDER, we are sympathetic with all of these replies. However, we think that we should accept these replies only on a consequentialist interpretation, and that when we do, we will see that there is a deeper reason for consequentialists to reject LOGGER; that is, even treating LOGGER as an open question is likely to be a net harm for nonhuman animals and other sentient beings in practice, for precisely the reasons that lead people to reject consequentialism.

Our aim in what follows, then, is to argue that, for a variety of reasons, a consequentialist criterion of rightness requires us to accept a partly nonconsequentialist decision procedure, and that this decision procedure conflicts with destroying animals and ecosystems at present (with certain caveats that we will explain). In particular, it requires us (p. 581) to place significant weight on protecting wild animal autonomy, cultivating virtuous character traits toward wild animals, and cultivating relationships of care with wild animals. This is centrally because exterminating animals negatively shapes our individual beliefs, values, and practices, and because living without wild animals altogether negatively shapes our collective beliefs, values, and practices. As earlier, in both cases, the result is that we tend to have attitudes that devalue animals and practices that harm them. How-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ever, we want to emphasize that we are less confident about how to evaluate LOGGER than about how to evaluate LARDER, for reasons that we will explain later.

4.2. The Individual Effects of Animal Extermination

We begin with the individual effects of animal extermination. We here focus on the individual effects of activities such as hunting, fishing, logging, and land development for human use, though we will consider later whether and to what degree these effects apply to other activities that reduce wild animal populations, too.

Our argument has two parts, which in many ways parallel the argument against LARDER.[50] First, we contend that participating in standard forms of extermination conditions humans to view animals as expendable, as inferior, and ultimately as having diminished moral status relative to humans. That is, when we performatively condone the killing of animals, directly or indirectly, in a way that treats these animals as mere means and undermines their agency, we condition ourselves to devalue and, as a result, harm other animals. Second, the more we open ourselves up to engaging in such practices in cases where they are a net benefit, the more willing we will be to support and engage in such practices in cases where they are not. Thus, we will argue, consequentialists should adopt a policy of not destroying animals and ecosystems by these means at all (with certain caveats that we will discuss).

Our first argument against LARDER focused centrally on two empirically validated social-psychological phenomena. First is the point that meat-eating creates psychological dissonance in people which they resolve by attributing lower mental life and moral status to nonhuman animals. Second is the closely related point that when people observe meat-eating, they infer that the people eating meat do not think that nonhuman animals are minded beings with moral standing. In our view, the best explanation for these findings is that people have at least partly deontological moral intuitions. If nonhuman animals have sentience and moral standing, they must be the kinds of beings who it is wrong to kill, eat, and exploit for human benefit. But since, the meat-eater judges, I and others *do* kill, eat, and exploit animals for human benefit, they must not have sentience and moral standing.

If many people have these kinds of moral intuitions, then we can predict that participating in the destruction of wild animals and their habitats will have similar consequences as participating in animal agriculture (again, covarying with the degree and (p. 582) kind of participation). That is, we can predict that this activity would cultivate within us an ideology of human supremacism (again, covarying with the degree and kind of participation). All of us have internalized deeply the idea that humans are the kinds of beings with whom we should have relationships of care, and that such relationships do not involve the kinds of violence and agency denial that is central to practices of hunting, fishing, and habitat destruction. Participating in these practices, then, creates differential psychological constructs regarding humans and other animals. Because we have also internalized the idea that building relationships of care with others is morally important, this may well lead us

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

to accept that our relationships with other animals are not as morally important as our relationships with other humans. Since these affective attitudes influence the decisions we make—from our recreational behavior to our voting behavior, political advocacy, career choice, philanthropic activity, conversations we have with others, and more—participating in the destruction of wild animals and their habitats negatively influences our individual and social treatment of nonhuman animals.

Brian Tomasik has argued explicitly against this kind of reasoning, urging us to help now and cultivate attitudes and relationships of care later.[51] Tomasik invites us to consider:

> [W]hat kinds of values are we trying to promote within society? Are we trying to promote the idea of holding back on doing the right thing because of how others may misinterpret it? … I think the ideology question isn't settled, because there's also value in challenging prevailing assumptions in the animal movement and promoting a culture of compassionate consequentialism, which could reduce the likelihood that the animal movement neglects huge sources of suffering in the future in the way it currently neglects … wild-animal suffering.

We agree with Tomasik that consequentialists should aim to cultivate and promote the virtues of responding with urgency and calculated efficiency to the suffering of nonhuman sentient beings. This may well require intervening to improve the welfare of wild animals sooner rather than later. But we nevertheless disagree with Tomasik on two significant points.

First, while consequentialists should cultivate virtues of urgency and efficiency, and while doing so sometimes conflicts with cultivating relationships of care, we believe that these activities are for the most part complementary. For example, if we aspire to respect wild animal life and autonomy while benefiting wild animals as much as possible within these constraints, such as by aiding them with medical intervention, reducing human and domesticated animal predation, and researching effective interventions into wild animal suffering, we can cultivate and promote anti-speciesist ideology and a concern for urgency and efficiency at the same time.

Second, consequentialists should be concerned about cultivating relationships of care with nonhuman animals not only because others are liable to misinterpret altruistically motivated extermination as speciesist, but also because we are liable to reinforce spe **(p. 583)** ciesism within ourselves and others whether or not we are misinterpreting our behavior as speciesist. The issue here is that participation in destroying animals and their environments would condition us to see them as having less sentience and moral standing independently of how we interpret our behavior. Granted, some interpretations might cause this effect to be larger than others. But we are suggesting that the effect would be present either way.

One aspect of our argument against LARDER focused on the observation that complex decision procedures have adjustable parameters that allow for false rationalization. We think that this consideration supports establishing deontological, virtue-theoretic, and

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

care-theoretic constraints on our utilitarian activity for domesticated animals and wild animals alike. In short, consequentialists should adopt decision procedures that pro tanto prohibit harming or killing nonhuman animals merely as means to further ends for much the same reason they should do so in the case of humans: the more we engage in such practices in anything other than clearly exceptional cases, the more willing we will be to engage in such practices in a wide range of cases that do not plausibly benefit wild animals.

In light of these considerations, we find it plausible that, even if an individual might be morally permitted to altruistically engage in wild animal extermination and habitat destruction in principle (i.e., in cases that idealize away facts about human psychology), most individuals are not morally permitted to take these actions in practice (i.e., in cases that do not idealize away facts about human psychology). Given the negative indirect effects of a policy of participating in the destruction of animals and habitats sometimes, a policy of not participating at all will do more good overall. Thus, ordinary consequentialists should instead adopt a policy of not participating at all, except perhaps in highly exceptional cases where doing so clearly does more good than harm.

With that said, we ultimately agree with Tomasik that these questions are unsettled. How we should resolve LOGGER will depend on our answers to many questions, especially questions about wild animal well-being and population ethics. Given how many wild animals there are, we are open to the possibility that the value of reducing their suffering via habitat destruction outweighs the value of reducing suffering more generally by cultivating virtues and relationships of care. For that reason, we are not claiming that LOGGER fails, but are rather claiming that it fails at present given our current epistemic state (which includes uncertainty about how much well-being wild animals have at present and could have in the future). On our best judgment, consequentialists should focus for now on helping wild animals in ways that respect their lives and autonomy, and on laying the groundwork for respectful, compassionate, and effective systematic interventions to reduce wild animal suffering in the future, as we will now discuss.

4.3. The Social Effects of Animal Extermination

We now consider the social effects of exterminating animals. (As earlier, we focus on the social effects of systems of animal extermination themselves, though we believe that (p. 584) individual support for these systems can have social effects, too.) In this case, we must consider not only the social effects of living in a world with legally sanctioned destruction of wild animals and habitats, but also the social effects of living in the world that this activity would bring about. Since the former effects are easier to infer from our earlier discussion than our analysis of the latter effects, we will focus on the latter effects here.

In particular, we will focus on three possible ways of structuring society: living with wild animals, living without wild animals (or at least, living with fewer wild animals) via domestication, and living without wild animals (or at least, living with fewer wild animals)

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

via extinction. Of course, in focusing on these options, we are not suggesting that they are exhaustive, since various combinations are possible as well. We suggest only that an initial focus on these options helps us to see clearly some of the relevant considerations.

Our argument has two parts. First, we contend that each alternative arrangement has its own ideological costs and benefits, significantly determining the possible relationships we could have with sentient beings in the future. Second, pursuing the best version of each arrangement is no guarantee that we will achieve that version, and we may instead be left with a warped version that looks more like a dystopian version of the status quo.

Consider first the effect that learning to live with wild animals might have. In the best case, we could learn to live with wild animals in a radical new way, respecting their lives and autonomy while intervening into their affairs to improve their well-being. This approach has the advantage of being more achievable than other approaches we will discuss. It would challenge human supremacism, producing an ideology of respect and compassion for sentient beings and teaching us lessons about coexistence and cooperation.[52] However, this approach would likely leave unaddressed some of the most significant sources of wild animal suffering, such as predation and r-selection.

Of course, there is a nontrivial chance that, if we choose to live with wild animals, we would not realize this best-case scenario. As Tomasik argues, it would be easy for us to slide back into our current state of indifference. In this case, we would neither improve the lives of wild animals nor challenge our current ideological presuppositions about wild animals and other sentient nonhuman beings. Our relationship with wild animals would continue to be one of mystery and awe, but also of alterity and indifference, characterized by the belief that wild animals should be left alone except where their human interests can be served by interfering with their lives. While learning to live with wild animals raises the quasi-utopian possibility of forming radical relationships of respect, compassion, coexistence, and assistance, it also raises the dystopian possibility of leaving the status quo forever intact.

Consider second the effect that learning to live without wild animals (or at least, living with fewer wild animals) via domestication might have. In the best case, we could domesticate wild animals by pursuing radical forms of sanctuary that look little like the current status quo for domesticated animals. Such forms of sanctuary would parentalistically give humans control over the forms of life wild animals could pursue, but would also be (p. 585) as deferential as possible to the revealed preferences of these animals. This system would provide wild animals with much higher levels of well-being on average, and it would also disrupt human supremacist ideology by teaching us lessons of care, responsibility, and stewardship. At the same time, it would be costly to develop and maintain, and it would risk reinforcing a diminished view of animal agency.

Again, there is a nontrivial chance that, if we choose to domesticate wild animals, we would not realize this best-case scenario. If advocates pursued the domestication of wild animals but without challenging our assumptions of human superiority, or if we continued to pursue conservation through our current frameworks, this could lead us to impose on

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

wild animals the status quo for animals living under human domestication, for example, confining wild animals in zoos. This system would provide domesticated wild animals with relatively low well-being, and it might also reinforce much the same ideology as zoos, teaching us "a false sense of our place in the natural order."[53] While domesticating wild animals raises the quasi-utopian possibility of forming radical relationships of care, responsibility, and stewardship, it also raises the dystopian possibility of imposing the current status quo for domesticated animals on a much higher proportion of sentient beings than we currently do.

Consider finally the effect that learning to live without wild animals (or at least, living with fewer wild animals) via extinction would have on human ideology. In the best case, we could bring about the extinction of wild animals through deliberate and cautious intervention that minimizes wild animal suffering and respects wild animal agency as much as possible. This would result in a world with little to no wild animal suffering. It may also teach us lessons of care for the suffering of sentient beings as well as lessons of caution about the hazards inherent in the very existence of sentient life. However, it also risks reinforcing the harmful idea that we should respond to the suffering of others (human and nonhuman alike) by seeking to control or eliminate the sufferers rather than by helping to reduce or eliminate their suffering.[54]

Once again, there is a nontrivial chance that, if we choose to bring about the extinction of wild animals, we would not realize this best-case scenario. For if advocates push for the extinction of wild animals without challenging our assumptions of human superiority, we could bring about the extinction of wild animals through the means that have come to be the status quo: incautiously destroying wild animal habitats through hunting, fishing, development, and more. This might still lead to a world with no wild animal suffering. However, it would also reinforce our ideology of human supremacism, teaching us that nonhuman animals are not deserving of the same kind of respect as human beings. Moreover, such a radically incautious process of total annihilation would leave a (p. 586) trail of immense suffering in its wake, with many wild animals dying slow and painful deaths of deprivation.

As we can see, all three of these possible futures carry costs and benefits, both directly (via our impact on wild animals) and indirectly (via our impact on human ideology). This is true for both the ideal and the nonideal versions of these possible futures.

It can be tempting to draw a strong conclusion on the basis of these considerations, but our view is that these considerations are far too preliminary to support such conclusions. After all, we remain highly uncertain about the experiences of wild animals, about the feasibility of each system, about the costs and opportunity costs of pursuing each system, and much more.

It can also be tempting not to draw a conclusion at all, instead urging caution until we have much more information. But we must remember that a precautionary approach is, in

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

practice, a choice to maintain a status quo that involves the continuing suffering of possibly septillions of sentient beings.

All things considered, our own weakly held view is that we should wait to take systematic action. If advocates invest resources in building capacity for research on reducing wild animal suffering and advocacy for the moral and political standing of wild animals, then we will likely be much better able to take informed and effective action in a few decades than we are now.[55] At present, we are not yet willing to take large-scale action for the sake of wild animals, and even if we were, we are not yet able to take such action without destabilizing the entire biosphere. Granted, playing the long game carries the cost of preserving the status quo in the short term. However, this cost is relatively minor compared to the epistemic and practical resources we can expect to gain through research and advocacy, given how few resources we have at the present time.

To be clear about the structure of our argument, what these social psychological considerations do is raise the moral costs of destroying wild animals and ecosystems. We think that these expected costs are sufficiently high that the expected benefit of eliminating negative well-being in wild animals does not outweigh them. Once we combine this consideration with the considerations that we discussed earlier (about our uncertainty about the total welfare of wild animals and the unpredictable consequences of intervention), the case for adopting a quasi-conservationist ethic becomes even stronger.

As earlier, this partly consequentialist, partly nonconsequentialist decision procedure is likely to result in at least some blameless wrongdoing. There might be some cases where destroying wild animals and habitats is best, and where we are not in a position to see that an exception is warranted. But again, this is fine. No decision procedure is perfect, and our suggestion is only that this partly consequentialist, partly nonconsequentialist decision procedure is best for most people in most situations at present.

We should stress that our argument is tentative. We are suggesting that LOGGER fails at present, given our current information state. Consequentialists have strong (in our view decisive) reason to reject LOGGER for now, and to instead accept principles which forbid destroying wild animals and ecosystems in all but the most exceptional cases. To **(p. 587)** be clear, we can imagine changing our minds with more information. For example, if we come to think that the aggregate well-being of wild animals is bad enough that the harm of allowing them to exist clearly outweighs the harm (both to wild animals and to other sentient beings) of cultivating and promoting human supremacist beliefs, values, and practices, we might come to think that LOGGER succeeds. However, we are currently skeptical that we will reach this conclusion.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

5. Conclusion: Future Technology, Future Directions

Anti-speciesist consequentialists and nonconsequentialists can agree that factory farming and wild animal welfare are two of the very highest-priority areas on which to spend scarce resources. However, many have supposed that consequentialists and nonconsequentialists are forced to disagree about the means of helping farmed animals and wild animals. Defenders of the Logic of the Larder have argued that consequentialism sometimes requires eating farmed animals in order to ensure that animals with positive well-being exist, while defenders of the Logic of the Logger have argued that consequentialism sometimes requires destroying wild animals and ecosystems in order to ensure that animals with negative well-being do not exist. In this chapter, we have argued that the Logic of the Larder and the Logic of the Logger both underestimate the importance of indirect decision procedures. In particular, they underestimate the role that our individual and collective policies play in shaping our moral attitudes and behavior and they underestimate the importance of accepting policies that are robust against harmful deviation. Once we have properly accounted for these considerations, it is clear that the Logic of the Larder fails and it is unclear that the Logic of the Logger succeeds.

We can expect future technological change to bring with it new and immense challenges for consequentialist moral theorists and advocates. Where the variety and number of farmed animals and wild animals have raised cluelessness and demandingness challenges for consequentialism, future sentient beings such as artificially intelligent minds will introduce even more varied and numerous minds into the world, thereby exacerbating these challenges even further. As a result, we can expect these advances to raise many new and difficult questions about the practical implications of consequentialism and about its deviation from nonconsequentialism. If we are wise, we will begin to develop and answer some of these questions now, before we have another moral tragedy on the scale of factory farming or wild animal suffering on our hands. It will be difficult to know, in advance, what kinds of future sentient beings might exist as technology continues to advance with increasingly accelerating returns, or when we will even recognize these sentient beings *as* sentient beings. But for precisely these reasons, we need to begin, now, to determine how consequentialism requires us to act in the face of such massive uncertainty, and we must work to identify, now, the indirect decision heuristics (p. 588) that will guide us away from moral dystopia before it arrives, rather than responding to it once it is already here.

References

Adams, C. J. 2015. *The Sexual Politics of Meat: A Feminist-Vegetarian Critical Theory*. New York: Bloomsbury.

Animal Charity Evaluators. 2017. “The Philosophical Foundation of Our Work.” **https://animalcharityevaluators.org/about/background/our-philosophy/**.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Arrhenius, G. 2012. “The Impossibility of a Satisfactory Population Ethics.” In *Descriptive and Normative Approaches to Human Behavior*, edited by E. Dzhafarov and P. Lacey, 1–26. Singapore: World Scientific.

Bastian, B., Loughnan, S., Haslam, N., and Radke, H. R. 2012. “Don’t Mind Meat? The Denial of Mind to Animals Used for Human Consumption.” *Personality and Social Psychology Bulletin* 38, no. 2: 247–256.

Beckstead, N. 2013. *On the Overwhelming Importance of Shaping the Far Future.* PhD diss., Rutgers University-Graduate School-New Brunswick.

Bentham, J. 1879. *An Introduction to the Principles of Morals and Legislation*. Oxford: Clarendon Press.

Bilz, K., and Nadler, J. 2009. “Law, Psychology, and Morality.” *Psychology of Learning and Motivation* 50: 101–131.

Brandt, R. B. 1984. “Utilitarianism and Moral Rights.” *Canadian Journal of Philosophy* 14, no. 1: 1–19.

Bratanova, B., Loughnan, S., and Bastian, B. 2011. “The Effect of Categorization as Food on the Perceived Moral Standing of Animals.” *Appetite* 57, no. 1: 193–196.

Broome, J. 2018. “Against Denialism.” *The Monist* 102, no. 1: 110–129.

Budolfson, M. 2015. “Is It Wrong to Eat Meat from Factory Farms? If So, Why?” In *The Moral Complexities of Eating Meat*, edited by B. Bramble and B. Fischer, 89–98. New York: Oxford University Press.

Buttlar, B., and Walther, E. 2019. “Dealing with the Meat Paradox: Threat Leads to Moral Disengagement from Meat Consumption.” *Appetite* 137: 73–80.

Cocking, D., and Oakley, J. 1995. “Indirect Consequentialism, Friendship, and the Problem of Alienation.” *Ethics* 106, no. 1: 86–111.

Cohen, Y. and Timmerman, T. 2016. “Actualism Has Control Issues.” *Journal of Ethics and Social Philosophy* 10, no. 3: 1–18.

Cowen, T. 2005. “Market Failure for the Treatment of Animals.” **http://www.gmu.edu/jbc/Tyler/animals.doc**.

Davis, A. Y. 2011. *Are Prisons Obsolete?* New York: Seven Stories Press.

Delon, N., and Purves, D. 2018. “Wild Animal Suffering Is Intractable.” *Journal of Agricultural and Environmental Ethics* 31, no. 2: 239–260.

Diamond, C. 1978. “Eating Meat and Eating People.” *Philosophy* 53, no. 206: 465–479.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Dickens, M. 2016. “Evaluation Frameworks, or When Importance/Neglectedness/Tractability Doesn’t Apply.” *Effective Altruism Forum*. **https://forum.effectivealtruism.org/posts/fMDxYL7ehWeptQ66r/evaluation-frameworks-or-when-importance-neglectedness**.

Donaldson, S., and Kymlicka, W. 2011. *Zoopolis: A Political Theory of Animal Rights*. New York: Oxford University Press.

(p. 589) Feldman, F. 1986. *Doing the Best We Can: An Essay in Informal Deontic Logic*. New York: Springer.

Fischer, B. 2019. *The Ethics of Eating Animals: Usually Bad, Sometimes Wrong, Often Permissible*. New York: Routledge.

Flores, A. R., and Barclay, S. 2015. *Trends in Public Support for Marriage for Same-Sex Couples by State*. Los Angeles: Williams Institute, UCLA School of Law.

Foucault, M. 1988. *Madness and Civilization: A History of Insanity in the Age of Reason*. New York: Vintage.

Frey, R. G. 1983. *Rights, Killing, and Suffering*. Oxford: Blackwell.

Greaves, H., and MacAskill, W. Unpublished manuscript. “The Case for Longtermism.”

Groff, Z., and Ng, Y. 2019. “Does Suffering Dominate Enjoyment in the Animal Kingdom? An Update to Welfare Biology.” *Biology & Philosophy* 34, no. 40: 1–16.

Gruen, L. 2011. *Ethics and Animals: An Introduction*. Cambridge: Cambridge University Press.

Gruen, L. 2014. “Dignity, Captivity, and an Ethics of Sight.” In *The Ethics of Captivity*, edited by L. Gruen, 231–247. London: Oxford University Press.

Gruen, L. 2015. *Entangled Empathy: An Alternative Ethic for Our Relationships with Animals*. Brooklyn: Lantern Books.

Gruen, L., and Jones, R. C. 2016. “Veganism as an Aspiration.” In *The Moral Complexities of Eating Meat*, edited by B. Bramble and B. Fischer, 153–171. New York: Oxford University Press.

Gustafsson, J. E. 2016. “Consequentialism with Wrongness Depending on the Difficulty of Doing Better.” *Thought: A Journal of Philosophy* 5, no. 2: 108–118.

Hanson, R. 2002. “Why Meat Is Moral, and Veggies Are Immoral.” **http://mason.gmu.edu/~rhanson/meat.html**

Hare, R. M. 1981. *Moral Thinking: Its Levels, Method, and Point*. New York: Oxford University Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Hare, R. M. 1993. “Why I Am Only a Demi-Vegetarian.” In *Essays on Bioethics*, edited by R. M. Hare, 233–246. London: Oxford University Press.

Hestermann, N., Le Yaouanq, Y., and Treich, N. Unpublished manuscript. “An Economic Model of the Meat Paradox.” **http://rationality-and-competition.de/wp-content/uploads/discussion_paper/164.pdf**

Jamieson, D. 1985. “Against Zoos.” *Environmental Ethics: Readings in Theory and Application* 5: 97–103.

Kagan, S. 2011. “Do I Make a Difference?” *Philosophy & Public Affairs* 39, no. 2: 105–141.

Korsgaard, C. M. 2018. *Fellow Creatures: Our Obligations to the Other Animals*. New York: Oxford University Press.

Kunst, J. R., and Hohle, S. M. 2016. “Meat Eaters by Dissociation: How We Present, Prepare and Talk about Meat Increases Willingness to Eat Meat by Reducing Empathy and Disgust.” *Appetite* 105: 758–774.

Loughnan, S., Haslam, N., and Bastian, B. 2010. “The Role of Meat Consumption in the Denial of Moral Status and Mind to Meat Animals.” *Appetite* 55, no. 1: 156–159.

MacAskill, W. 2015. *Doing Good Better: Effective Altruism and a Radical New Way to Make a Difference*. London: Guardian Faber.

MacAskill, W. 2017. “Effective Altruism: Introduction.” *Essays in Philosophy* 18, no. 1: 1–5.

Matheny, G., and Chan, K. M. 2005. “Human Diets and Animal Welfare: The Illogic of the Larder.” *Journal of Agricultural and Environmental Ethics* 18, no. 6: 579–594.

McNaughton, D. 1998. “Consequentialism.” *Routledge Encyclopedia of Philosophy* 2: 603–606.

Mitchell-Brody, M., and Sebo, J. Unpublished manuscript. “Wildness and Civilization.”

(p. 590) Ng, Y. K. 1995. “Towards Welfare Biology: Evolutionary Economics of Animal Consciousness and Suffering.” *Biology and Philosophy* 10, no. 3: 255–285.

Parfit, D. 1984. *Reasons and Persons*. New York: Oxford University Press.

Piazza, J., Ruby, M. B., Loughnan, S., Luong, M., Kulik, J., Watkins, H. M., and Seigerman, M. 2015. “Rationalizing Meat Consumption. The 4Ns.” *Appetite* 91: 114–128.

Portmore, D. W. 2009. “Consequentializing.” *Philosophy Compass* 4, no. 2: 329–347.

Posner, R. A. 2004. “Animal Rights: Legal, Philosophical, and Pragmatic Perspectives.” In *Animal Rights: Current Debates and New Directions*, edited by C. R. Sunstein and M.C. Nussbaum, 51–77. New York: Oxford University Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Reese, J. 2019. “US Factory Farming Estimates.” **https://www.sentienceinstitute.org/us-factory-farming-estimates**

Rothgerber, H. 2015. “Can You Have Your Meat and Eat It Too? Conscientious Omnivores, Vegetarians, and Adherence to Diet.” *Appetite* 84: 196–203.

Rowlands, M. 1997. “Contractarianism and Animal Rights.” *Journal of Applied Philosophy* 14, no. 3: 235–247.

Salt, H. S. 1914. *The Humanities of Diet: Some Reasonings and Rhymings*. Manchester: Vegetarian Society.

Schlottmann, C., and Sebo, J. 2018. *Food, Animals, and the Environment: An Ethical Approach*. New York: Routledge.

Schubert, S., and Garfinkel, B. 2017. “Hard-to-Reverse Decisions Destroy Option Value.” *Centre for Effective Altruism*. **https://www.centreforeffectivealtruism.org/blog/hard-to-reverse-decisions-destroy-option-value/**

Sebo, J., and Singer, P. 2018. “Activism.” In *Critical Terms for Animal Studies*, edited by L. Gruen, 33–46. Chicago, IL: University of Chicago Press.

Sidgwick, H. 1874. *The Methods of Ethics*. London: Macmillan; reprinted 1877, 1884, 1890, 1893, 1901, 1907.

Singer, P. 1999. “A Response.” In *Singer and His Critics*, edited by D. Jamieson, 269–335. Malden, MA: Blackwell.

Singer, P. 2011. *Practical Ethics.* 3rd ed. New York: Cambridge University Press.

Sinhababu, N. 2018. “Scalar Consequentialism the Right Way.” *Philosophical Studies* 175, no. 12: 3131–3144.

Sinnott-Armstrong, W. 2019. “Consequentialism.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. **https://plato.stanford.edu/archives/sum2019/entries/consequentialism/**>.

Stephen, L. 1896. *Social Rights and Duties: Addresses to Ethical Societies*. New York: Macmillan.

Tankard, M. E., and Paluck, E. L. 2016. “Norm Perception as a Vehicle for Social Change.” *Social Issues and Policy Review* 10, no. 1: 181–211.

Tankard, M. E., and Paluck, E. L. 2017. “The Effect of a Supreme Court Decision Regarding Gay Marriage on Social Norms and Personal Attitudes.” *Psychological Science* 28, no. 9: 1334–1344.

Taylor, S. 2017. *Beasts of Burden: Animal and Disability Liberation.* New York: The New Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Tomasik, B. 2016. “Efforts to Help Wild Animals Should Be Effective, Not Idealistic.” *Essays on Reducing Suffering*. **https://reducing-suffering.org/habitat-loss-not-preservation-generally-reduces-wild-animal-suffering/**.

Tomasik, B. 2017a. “Habitat Loss Generally Reduces Wild-Animal Suffering.” *Essays on Reducing Suffering*. **https://reducing-suffering.org/habitat-loss-not-preservation-generally-reduces-wild-animal-suffering/**.

(p. 591) Tomasik, B. 2017b. “I’m Not a Speciesist; Just a Utilitarian.” *Essays on Reducing Suffering*. **https://reducing-suffering.org/im-not-speciesist-im-just-utilitarian/**.

Tomasik, B. 2018. “How Many Wild Animals Are There?” *Essays on Reducing Suffering*. **https://reducing-suffering.org/how-many-wild-animals-are-there/**.

Torres, R. 2015. “Altruistic Murders.” *Contragaia*. **https://contragaia.wordpress.com/2015/10/19/altruistic-murders/**.

Wiblin, R. 2016. “The Important/Neglected/Tractable Framework Needs To Be Applied with Care.” *Effective Altruism Forum*. **https://forum.effectivealtruism.org/posts/74oJS32C6CZRC4Zp5/the-important-neglected-tractable-framework-needs-to-be**.

Wiland, E. 2007. “How Indirect Can Indirect Utilitarianism Be?” *Philosophy and Phenomenological Research* 74, no. 2: 275–301.

Notes:

([1]) This paper benefited from feedback from and discussion with Mark Budolfson, Tomi Francis, Maryse Mitchell-Brody, Doug Portmore, Abraham Rowe, Alexa Stonebarger, Travis Timmerman, and members of the 2018 Global Priorities Institute summer works in progress group. We have also benefited from countless conversations with many animal advocates over the years.

([2]) Animal Charity Evaluators (2017).

([3]) Feldman (1986); Hare (1981); Parfit (1984, 24–28); Sidgwick (1874, 489–490).

([4]) We take maximizing consequentialism to be compatible with certain forms of scalar consequentialism, such as those developed by Gustafsson (2016) and Sinhababu (2018), in that all such views regard maximizing the good as uniquely maximally right.

([5]) Portmore (2009); Sinnott-Armstrong (2019).

([6]) Bentham (1879).

([7]) For examples, see Donaldson and Kymlicka (2011), Korsgaard (2018), and Rowlands (1997).

([8]) MacAskill (2017).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([9]) In a 2017 survey, about two-thirds of EAs reported accepting or leaning toward consequentialism.

([10]) MacAskill (2015).

([11]) Dickens (2016); Wiblin (2016).

([12]) Schlottmann and Sebo (2018).

([13]) Tomasik (2018).

([14]) Delon and Purves (2018).

([15]) Hare (1981).

([16]) Brandt (1984); Cocking and Oakley (1995); Hare (1981); McNaughton (1998); Sidgwick (1874); Wiland (2007).

([17]) Sebo and Singer (2018).

([18]) CAFOs are defined by the EPA as farms with upward of 37,500 meat chickens and upward of 25,000 laying hens, respectively (Reese 2019).

([19]) Gruen (2011, 83).

([20]) Hanson (2002). See also Cowen (2005); Hare (1993); Posner (2004); Salt (1914); Singer (1999); and Stephen (1896).

([21]) The question whether eating meat in fact increases the demand for meat and so causes future animals to exist is one that has been thoroughly explored elsewhere, and we do not take it up here. For our part, we find persuasive Kagan's (2011) reasoning. For a persuasive parallel discussion in the context of climate change, see Broome (2018). See also Budolfson (2015); Gruen and Jones (2016); Schlottmann and Sebo (2018); and Singer (2011).

([22]) For more on the Mere Addition Principle, see Arrhenius (2012).

([23]) Donaldson and Kymlicka (2011, chaps. 2 and 4); Gruen (2011, chap. 3).

([24]) Matheny and Chan (2005).

([25]) As we will note, others have explored this option, too, including Fischer (2019), Gruen (2011), and Singer (2011).

([26]) Both of these arguments assume actualism rather than possibilism about obligation logic. Actualism is the view that I ought to take an action if and only if what will happen if I take that action is better than what will happen if I do not take that action, whereas possibilism is the view that I ought to take an action if and only if taking that action is part of the best maximally-specific act-set that I can perform. Whether possibilists should accept our arguments depends a great deal on the details of their individual views. For more on

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

this distinction, see: Cohen and Timmerman (2016) and Cohen and Timmerman in this volume.

([27]) Singer (2011, 134).

([28]) Diamond (1978, 467).

([29]) Gruen (2011, 103).

([30]) Adams (2015, 100).

([31]) Bastian et al. (2012, 249–250).

([32]) Loughnan, Haslam, and Bastian (2010).

([33]) Bratanova, Loughnan, and Bastian (2011).

([34]) Kunst and Hohle (2016).

([35]) Buttlar and Walther (2019).

([36]) Rothgerber (2015).

([37]) For discussion of the social effects of individual support for systems of animal exploitation, see Schlottmann and Sebo (2018, chap. 9).

([38]) Cf. Piazza et al. (2015); Gruen (2014).

([39]) Davis (2011, 18–19).

([40]) Davis (2011, 17).

([41]) Davis (2011, 22–39).

([42]) Bilz and Nadler (2009); Flores and Barclay (2015); Tankard and Paluck (2016); Tankard and Paluck (2017).

([43]) Bilz and Nadler (2009), 104.

([44]) For more on the overwhelming importance of shaping the far future for determining right action for consequentialists, see Beckstead (2013) and Greaves and MacAskill (unpublished manuscript).

([45]) Tomasik (2017a).

([46]) Ng (1995).

([47]) Donaldson and Kymlicka (2011); Singer (2011); Tomasik (2017a).

([48]) Groff and Ng (2019).

([49]) Delon and Purves (2018).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([50]) As with our arguments in Section 3.2, our arguments in Section 4.2 assume actualism.

([51]) Tomasik (2016). Cf. Tomasik (2017b).

([52]) Gruen (2015).

([53]) Jamieson (1985).

([54]) This oppressive idea can harm humans as well. For example, in cases where people with mental and physical difference are suffering, many people see this suffering as a reason to reduce mental and physical difference in the world, rather than as a reason to create a world that can accommodate mental and physical difference. For more on this subject, see Foucault (1988), Mitchell-Brody and Sebo (unpublished manuscript), and Taylor (2017).

([55]) Cf. Schubert and Garfinkel (2017).

Tyler M. John

Tyler M. John is a PhD student in philosophy at Rutgers University–New Brunswick. His main areas of research are distributive ethics, political philosophy of the long-term future, and animal moral, legal, and political philosophy. He is a coauthor of Chimpanzee Rights: The Philosophers' Brief (2018) and of articles appearing in Ethics and Economics and Philosophy.

Jeff Sebo

Jeff Sebo is Clinical Associate Professor of Environmental Studies, Affiliated Professor of Bioethics, Medical Ethics, and Philosophy, and Director of the Animal Studies M.A. Program at New York University. He works primarily on bioethics, animal ethics, and environmental ethics. His coauthored books Chimpanzee Rights and Food, Animals, and the Environment are currently available from Routledge, and his book Why Animals Matter for Climate Change is currently in contract with Oxford University Press. Jeff is also on the Board of Directors at Animal Charity Evaluators, the Board of Directors at Minding Animals International, and the Executive Committee at the Animals & Society Institute.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Introduction

Douglas W. Portmore

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy, Moral Philosophy
Online Publication Date: Oct 2020 DOI: 10.1093/oxfordhb/9780190905323.013.33

Abstract and Keywords

I argue that a theory is consequentialist if and only if it is, in the important respects, sufficiently similar to classical utilitarianism. Unfortunately, though, philosophers can't seem to agree on what the important respects are. So there is no one way that the term "consequentialism" is used, but only several different ways that it's used by various sets of philosophers with different views about what's most important about classical utilitarianism. But if, like many philosophers, we accept that what's most important about classical utilitarianism is that it takes the deontic statuses of actions to be a function of how various possible outcomes rank, then we can, I show, reconcile consequentialism with deontology. And I explore whether this ability to be reconciled with other theories, such as deontology, undermines the importance of the consequentialism/nonconsequentialism distinction. I then end by summarizing each of this anthology's four parts and the issues that are explored in their corresponding chapters.

Keywords: agent-relative, consequentialism, constraints, Kantianism, deontology, utilitarianism

THE word "consequentialism" is a term of art, and different philosophers use it to mean different things. This makes it difficult to define—an issue that I'll return to in the next section. But, perhaps, philosophers would agree to this much: a theory is consequentialist only if it holds that the deontic statuses of actions (or the reasons to perform them) are some function of how various possible outcomes rank.[1] What's more, most consequentialists go further and hold that this ranking is explanatorily prior both to the deontic statuses of actions and to the reasons for performing them. For instance, traditional act consequentialists hold not only that an act is permissible if and only if it maximizes the good, but also that what makes an act permissible is that it maximizes the good.[2] More precisely, they accept the following.

> *Traditional Act Consequentialism (TAC)*: For any subject S and any available act φ: ($TAC_{P\text{-}Specifying}$) S's φ-ing is morally permissible if and only if there is no available alternative act ψ whose outcome is better than that of her φ-ing. And ($TAC_{P\text{-}Making}$)

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

> the most fundamental permissibility-making feature of a permissible act is its lacking an alternative whose outcome is better than its own.[3]

(p. 2) According to TAC, I'm permitted to type this sentence if and only if there is no available alternative (such as my going for a run or typing out a different sentence) whose outcome is better than that of my typing this sentence. And, as contemporary consequentialists understand the notion, the *outcome* of an event (such as my typing this sentence) includes not only its causal consequences (such as the appearance of this sentence on my computer's monitor), but everything that would be the case given its occurrence.[4] Thus, the fact that someone with the surname "Portmore" typed the word "surname" on July 11, 2019, is part of the outcome of my typing this sentence. And the fact that you're wearing whatever it is that you're wearing as you now read this is also part of the outcome of my having typed that sentence on July 11, 2019. More precisely, then, the outcome of an event is the maximally specific way that the world would be if it were to occur.

Now, there's a way that the world would be if I were to type this sentence as well as a way that the world would be if I were to perform some available alternative.[5] And TAC holds that my typing this sentence is permissible if and only if there is no available alternative such that the way the world would be if I were to perform this alternative is better than the way the world would be if I were to type this sentence. Moreover, given its commitment to $TAC_{P\text{-}Making}$, TAC holds that, if it's permissible for me type this sentence, it's in virtue of the fact that there's no alternative whose outcome is better than that of my typing this sentence.

So, given its commitment to $TAC_{P\text{-}Making}$, the following seems not to be a version of TAC, for it seems to provide a different account of what the most fundamental permissibility-making feature of a permissible act is.

> *Kantsequentialism (KQ)*: For any subject S and any available act φ: ($KQ_{P\text{-}Specifying}$) S's φ-ing is morally permissible if and only if S's φ-ing accords with Kant's categorical imperative. ($KQ_{P\text{-}Making}$) The most fundamental permissibility-making feature of a permissible act is its being in accordance with Kant's categorical imperative. And ($KQ_{CI\text{-}Specifying}$) S's φ-ing accords with Kant's categorical imperative if and only if there is no available alternative act ψ whose outcome is better than that of her φ-ing.

Yet, since $KQ_{P\text{-}Specifying}$ and $KQ_{CI\text{-}Specifying}$ jointly entail $TAC_{P\text{-}Specifying}$, Kantsequentialism yields the exact same set of deontic verdicts that TAC does. Thus, on both TAC and Kantsequentialism, the properties of being morally permissible and of maximizing the good are necessarily coinstantiated. Nevertheless, it appears that Kantsequentialism contradicts TAC concerning what ultimately makes an act permissible. For whereas TAC (given its commitment to $TAC_{P\text{-}Making}$) holds that what ultimately makes an act (p. 3) permissible is that it maximizes the good, Kantsequentialism (given its commitment to $KQ_{P\text{-}Making}$) holds that what ultimately makes an act permissible is that it accords with Kant's categorical imperative.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Nevertheless, appearances can be deceiving. For instance, it may seem that if one person claims that the boiling point of water is 100° Celsius and another claims that it is 212° Fahrenheit, they must be disagreeing. But, in fact, they agree on what the boiling point of water is, as they're just using two different notational systems to refer to the exact same temperature. After all, 100° Celsius just is 212° Fahrenheit. So each is making the exact same claim, only using different terminology. And some philosophers believe that the same holds for the putative difference between claims like $TAC_{P\text{-}Making}$ and $KQ_{P\text{-}Making}$. They hold that these two seemingly distinct claims are really just the exact same claim stated in different terms. For they hold that the property of lacking an alternative whose outcome is better than its own just is the property of being in accordance with Kant's categorical imperative in exactly the same way that the property of being 100° Celsius just is the property of being 212° Fahrenheit. What's more, some of these same philosophers believe that every plausible seemingly nonconsequentialist theory has a consequentialist counterpart that's extensionally equivalent to it, just as Kantsequentialism has TAC as its extensionally equivalent consequentialist counterpart. And at least one of these philosophers—specifically, Jamie Dreier (2011)—goes so far as to conclude from this that every plausible seemingly nonconsequentialist theory is in fact a consequentialist theory. For he holds that such seemingly nonconsequentialist theories are in fact mere notational variants on their extensionally equivalent consequentialist counterparts.

I won't try to assess Dreier's claims here, for my point is only that philosophers disagree on whether Kantsequentialism is consequentialist.[6] As we've just seen, some (e.g., Dreier) think that Kantsequentialism and TAC are mere notational variants on the exact same theory. Consequently, they hold that both theories are consequentialist (and also that both are Kantsequentialist). But others deny that they're the same theory. And these others divide into two types. One type (e.g., Portmore 2011) holds that, whereas TAC is consequentialist, Kantsequentialism is not. And the other type (e.g., Brown 2011) holds that, although TAC and Kantsequentialism are distinct theories, they're both consequentialist—just two distinct forms of consequentialism. These others hold that a theory is consequentialist if and only if it takes the properties of being morally permissible and of maximizing the good to be necessarily coinstantiated (see Brown 2011, 753). And, since both TAC and Kantsequentialism hold that these properties are necessarily coinstantiated, they both count as consequentialist. So one thing that philosophers disagree on is whether extensionally equivalent theories such as TAC and Kantsequentialism are distinct theories. And another thing that they disagree on is whether consequentialists are committed, not only to the coinstantiation of properties such as being morally permissible and maximizing the good, but also to the latter's having explanatory priority over the former. That is, they disagree on whether consequentialists are committed to only $TAC_{P\text{-}Specifying}$ or to both $TAC_{P\text{-}Specifying}$ and $TAC_{P\text{-}Making}$.

(p. 4) Of course, there's a lot that philosophers can agree on concerning consequentialism. For one, philosophers seem to agree that consequentialists needn't take the relevant outcomes to be those of the acts themselves. The relevant outcomes can instead be those resulting from something related to the acts, such as the formation of a motive that would incline one to perform such acts or the acceptance of a code of rules that requires one to

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

perform such acts. Thus, most philosophers agree that the following is a version of consequentialism.[7]

> *Rule Consequentialism (RC)*: For any subject S and any available act φ: ($RC_{P\text{-}Specifying}$) S's φ-ing is morally permissible if and only if it is permitted by the ideal code, where the ideal code is, say, the code of rules whose acceptance by the vast majority of everyone everywhere has greater expected goodness than that of any alternative code.[8] And ($RC_{P\text{-}Making}$) the most fundamental permissibility-making feature of a permissible act is its being permitted by the ideal code.

So, whereas act consequentialism holds that what makes an act obligatory is that its outcome outranks that of every *alternative act*, rule consequentialism holds that what makes an act obligatory is that it's required by the code of rules whose associated outcome outranks that of every *alternative code*.

For another, philosophers seem to agree that consequentialists needn't give any account of the deontic statuses of actions. They agree, that is, that consequentialists may account only for our reasons for action, taking them to be a function of how the outcomes of those acts rank. Thus, the following is generally recognized as being a version of consequentialism.

> *Scalar Consequentialism (SC)*: For any subject S and any available act φ: ($SC_{R\text{-}Specifying}$) The better the outcome of S's φ-ing the more moral reason S has to φ. And ($SC_{R\text{-}Providing}$) the most fundamental reason-providing feature of an act is its having a good outcome.[9]

1. Trying to Define Consequentialism

As we've seen, a necessary condition for a theory's being consequentialist is that it holds that the deontic statuses of actions or the reasons to perform them are some function of (p. 5) a ranking of various possible outcomes. Yet philosophers disagree on whether this is also sufficient. For whereas some philosophers insist that a theory counts as consequentialist only if its ranking is agent-neutral, others allow that any ranking (agent-relative or agent-neutral) will do. Thus, philosophers disagree on whether the following is a version of consequentialism.

> *Ethical Egoism (EE)*: For any subject S and any available act φ: ($EE_{P\text{-}Specifying}$) S's φ-ing is morally permissible if and only if there is no available alternative act ψ whose outcome is better for her than that of her φ-ing. And ($EE_{P\text{-}Making}$) the most fundamental permissibility-making feature of a permissible act is its lacking an alternative whose outcome is better for the agent than its own.

A ranking of outcomes is agent-relative if and only if we get different rankings for agents with different features. Thus, ethical egoism's ranking of outcomes is agent-relative, because it holds that whether one outcome outranks another depends on whether the one is

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

better *for the given agent* than the other, and that, in turn, depends on features of that agent—for example, on whether she finds this or that experience pleasing. More precisely, the distinction between agent-relative and agent-neutral rankings is as follows. Take any ranking of the following form: "the outcome of S's φ-ing outranks that of S's ψ-ing if and only if the outcome of S's φ-ing is [INSERT SOME PHRASE HERE] that of S's ψ-ing." A ranking of this form is agent-relative if and only if what replaces the bracketed phrase makes an essential reference to S, and it's agent-neutral otherwise. So, if what replaces it is something such as "better for S than" or "preferred by S to," then the ranking is agent-relative. But if what replaces it is something such as "better than" or "preferable to," then the ranking is agent-neutral. So, whereas ethical egoism's ranking, which appeals to "better for S than," is agent-relative, TAC's ranking, which appeals to "better than," is agent-neutral. And whereas some philosophers count both TAC and ethical egoism as consequentialist, those who insist that a theory counts as consequentialist only if its ranking of outcomes is agent-neutral deny that ethical egoism is consequentialist.

Why do philosophers disagree on whether theories such as Kantsequentialism and ethical egoism count as consequentialist? I believe that Walter Sinnott-Armstrong (2015) gives the correct diagnosis:

> In actual usage, the term 'consequentialism' seems to be used as a family resemblance term to refer to any descendant of classic[al] utilitarianism that remains close enough to its ancestor in the important respects. Of course, different philosophers see different respects as the important ones. Hence, there is no agreement on which theories count as consequentialist under this definition.

So a theory is consequentialist if and only if it is, in the important respects, close enough to the following archetypal form of consequentialism.

> *Classical Utilitarianism (CU)*: For any subject S and any available act φ: ($CU_{P\text{-}Specifying}$) S's φ-ing is morally permissible if and only if there is no available alternative act ψ (p. 6) whose outcome is better than that of her φ-ing. ($CU_{P\text{-}Making}$) The most fundamental permissibility-making feature of a permissible act is its lacking an alternative whose outcome is better than its own. And ($CU_{B\text{-}Specifying}$) there's no available alternative act ψ whose outcome is better than that of her φ-ing if and only if there's no available alternative act ψ whose hedonic utility is greater than that of her φ-ing.[10]

Of course, philosophers disagree on what the important respects are, and this is why they disagree on what counts as a consequentialist theory. Some think that what's important about classical utilitarianism is that it holds that the properties of being morally permissible and of maximizing the good are necessarily coinstantiated. And so they hold that any theory that takes these properties to be necessarily coinstantiated counts as consequentialist. Thus, Kantsequentialism counts as consequentialist on their view. But others think that what's important about classical utilitarianism is that it holds that what ultimately makes an act permissible is that it maximizes the good. On this view, Kantsequentialism doesn't count as consequentialist, because it instead holds that what ultimately makes an

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

act permissible is that it accords with Kant's categorical imperative—and I'm assuming now, merely for the sake of argument, that the claim that an act accords with Kant's categorical imperative isn't a mere notational variant on the claim that it maximizes the good.

Still others hold that what's important about classical utilitarianism is that it takes the deontic statuses of acts to be a function of some *agent-neutral* ranking of outcomes. So they deny that ethical egoism is consequentialist. Yet others think that what's important about classical utilitarianism is that it takes the deontic statuses of acts to be a function of some (not necessarily agent-neutral) ranking of outcomes, and so they allow that ethical egoism counts as consequentialist.

Of course, we might think that none of these views are right. For we might think instead that what's most important about classical utilitarianism is that its only fundamental commitment is to making the world as good as possible. After all, isn't this what most utilitarians take to be our ultimate moral goal? Interestingly, though, there are at least three reasons to question this account of what's important about classical utilitarianism. First, there are some examples that suggest that classical utilitarians can't take on this commitment. Second, even if they could, they may not want to, given that doing so makes their view subject to a potentially devastating objection. And, third, insofar as we both take rule consequentialism to be consequentialist and insist that a theory is consequentialist if and only if it resembles classical utilitarianism in the important respects, we must deny that what's important about classical utilitarianism is that its only fundamental commitment is to making the world as good as possible. For rule (p. 7) consequentialism must disavow such a commitment in order to avoid being incoherent. Let's consider each of these in turn.

First, the following example, which is borrowed with modifications from David Estlund (2017, 53), suggests that classical utilitarianism can't take on a commitment to making the world as good as possible.[11]

> ***Slice and Patch Go Golfing***: Unless a patient's tumor is removed this afternoon, he'll die (though not painfully). Cutting and stitching this afternoon by the only two available doctors, Slice and Patch, is the only thing that can save him. For Slice is the only one who can do the cutting, and Patch is the only one who can do the stitching. If there is either cutting without stitching or stitching without cutting, the patient's death will be physically agonizing. It would even be cruel for one of them to show up knowing that the other won't, as this would only needlessly get the patient's hopes up, making his death psychologically agonizing. Unfortunately, both Slice and Patch are bad people who want their patient to die. Consequently, each has freely decided to go golfing regardless of what she thinks the other might do, and there's no dissuading either of them. What's more, each knows this about the other. Thus, each knows that, given the other's obstinate unwillingness to do her part in saving the patient, going golfing is her only option for maximizing hedonic utility. So, in the end, each doctor enjoys a pleasant round of

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

golf while her patient dies. Thus, of the four collectively available outcomes, they together produce the one I've labeled O_4. See Table 1.1.

Table 1 Slice and Patch Go Golfing

	Slice cuts	**Slice goes golfing**
Patch stitches	(O_1) This is the best world that they could together produce: their patient lives, although neither doctor enjoys a pleasant round of golf.	(O_2) This is tied for the worst world that they could together produce: their patient dies in agony, but at least Slice enjoys a pleasant round of golf.
Patch goes golfing	(O_3) This is tied for the worst world that they could together produce: their patient dies in agony, but at least Patch enjoys a pleasant round of golf.	(O_4) This is the second-best world that they could together produce: their patient dies painlessly while both doctors enjoy a pleasant round of golf.

In this case, the patient dies even though the two doctors could have saved him. Thus, they fail to produce the best world that they could together produce. Even so, classical utilitarianism holds that neither of them is guilty of a moral violation. For, on classical utilitarianism, their only moral obligation is to maximize hedonic utility, and they do this so long as they together produce either O_1 or O_4. Unfortunately, there is, on classical utilitarianism, no duty for them to work together to make the world as good as possible by producing O_1 instead of O_4. Thus, classical utilitarianism doesn't include a fundamental commitment to making the world as good as possible.[12] And it's because of this that several consequentialists have rejected it.[13]

(p. 8) But even if classical utilitarians could take on such a commitment, they probably shouldn't. This is because such a commitment would render their theory subject to a potentially devastating objection. Let me explain.[14] To start, a fundamental commitment to making the world as good as possible implies that the only thing that we should fundamentally care about is the world and the extent of its goodness. And let's call this *world-only fundamentalism*. World-only fundamentalism implies that we should care about people only insofar as they're receptacles for things that make the world better—for example, pleasure. That is, this view implies that people matter only derivatively in that increasing their pleasure is a means to increasing the value of the world. And this seems misguided. For it seems that at least some of the things that we should ultimately care about are people and their interests. And, thus, we should care about increasing people's pleasure for the sake of those people, and not (or, at least, not just) for the sake of making the world better. To illustrate, assume, just for the sake of argument, that it's always

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

in someone's interest to experience some episode of pleasure but that this episode of pleasure will contribute to the value of the world only if it's warranted—that is, only if the thing that the subject is taking pleasure in is something that merits being the object of pleasure.[15] And now imagine that by φ-ing we could provide a man named Sonum with some unwarranted pleasure (say, the pleasure that he takes in listening to clamorous noises). Now, since the world-only fundamentalist cares only for the value of the world, which would not be increased by our φ-ing, she must insist that we have absolutely no reason to φ. But given that Sonum is a person and that people and their interests matter fundamentally, we should instead think that we have at least some reason to φ in that φ-ing would promote Sonum's interests even if it doesn't add to the value of the world.

Of course, classical utilitarians can avoid this worry, because it's not classical utilitarianism, but the view that our sole and ultimate moral goal is to make the world as good as possible, that entails world-only fundamentalism. So there's no reason why classical (p. 9) utilitarians can't hold that people and their interests ultimately matter and that, consequently, we should care about promoting hedonic utility not only because it's a means to promoting the value of the world but also because it's a means to promoting people's interests. Thus, the classical utilitarian need only disavow the view that our only fundamental commitment is to making the world as good as possible to avoid this potentially devastating objection.

Third, we should deny that what's important about classical utilitarianism is that its only fundamental commitment is to making the world as good as possible if we think both that rule consequentialism is consequentialist and that a theory is consequentialist if and only if it resembles classical utilitarianism in the important respects. For, as I'll now show, rule consequentialism must disavow a fundamental commitment to making the world as good as possible if it's to avoid incoherence. To see this, note that it's incoherent to hold both that we should comply with the ideal code only as a means to the ultimate goal of making the world as good as possible and that agents should comply with the ideal code even when it wouldn't be a means to making the world as good as possible. Therefore, if our sole and ultimate moral goal is to make the world as good as possible, then agents should not comply with the ideal code when this would knowingly thwart that goal. But, of course, if rule consequentialism is to avoid collapsing into act consequentialism, it must hold that the ideal code sometimes requires us to do something that would thwart the goal of making the world as good as possible. And this means that the only way for the rule consequentialist to avoid the incoherence objection without collapsing into act consequentialism is to disavow the view that our sole and ultimate moral goal is to make the world as good as possible.

This is why rule consequentialists such as Brad Hooker deny that rule consequentialism has an overarching commitment to maximizing the good. Indeed, Hooker believes that what ultimately matters is not that we make the world as good as possible, but that we each "behave in ways that are impartially defensible" (2000, 102).[16] But, of course, if Hooker holds that what ultimately matters is that we each behave in ways that are impartially defensible rather than in ways that will make the world as good as possible, we

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

might suspect that he is actually a crypto-contractualist as opposed to genuine consequentialist (2000, 104). Yet Hooker can use the term “consequentialism” anyway he pleases, so long as he’s clear on how he’s using it. And he is (2000, 104). So it seems that Hooker thinks that making the world as good as possible is not the only thing that ultimately matters and that what also matters are people and our showing adequate respect for them by acting in only those ways that are impartially defensible.

To sum up, we’ve seen that there is no illuminating set of necessary and sufficient conditions for a theory’s being consequentialist. At best, we can say that a theory is consequentialist if and only if it resembles classical utilitarianism in the important respects. (p. 10) But this is illuminating only if we have an account of what the important respects are. And, unfortunately, it’s impossible to give any clear and uncontroversial account of them given the extent to which philosophers disagree on this issue. This means that there is no one way that philosophers use the term “consequentialism,” but only several different ways that it’s used depending on what each group thinks is most important about classical utilitarianism.

2. The Importance of Consequentialism

Despite there being no consensus on how to define “consequentialism,” the distinction between consequentialism and nonconsequentialism has played a prominent role in contemporary Western philosophy.[17] Whether this is merited depends on how we are to understand the distinction. If, as Dreier believes, it’s like the distinction between Celsius and Fahrenheit in being merely a difference in notation, then it’s entirely unwarranted. And this is why Dreier (1993) has urged normative theorists to focus instead on more substantive distinctions, such as that between agent-relative theories and agent-neutral theories. If, however, we understand consequentialism to be a theory that is, by definition, agent-neutral, then we can easily make sense of its prominent role.[18] For, in that case, the consequentialism/nonconsequentialism distinction marks a major fault line in normative ethics: that between theories that can and theories that cannot accommodate the various agent-relative features of common-sense morality: such as special obligations, agent-centered options, and agent-centered constraints.[19]

Alternatively, we might understand consequentialism such that it lies between these two extremes, where it needn’t be agent-neutral but is much more than just a different notational system for describing moral theories. And if this is the way we understand consequentialism, then the significance of the consequentialism/nonconsequentialism distinction will lie, not with what sorts of features it can accommodate, but with what sort of explanation it gives for those features. To illustrate, let’s consider agent-centered constraints. There is, on a given theory, an agent-centered constraint against a subject’s performing an act of a certain type if and only if there is, on that theory, some possible set of circumstances in which it would be impermissible for her to perform an act of that type even though doing so would both minimize the total instances of actions of that type and have no other morally relevant implications (Scheffler 1985, 409). (p. 11) Thus, common-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

sense morality includes an agent-centered constraint (hereafter, simply "constraint") against murder (that is, the premeditated killing of an innocent person without her autonomous consent), given that it prohibits a subject from committing murder even to prevent two others from each committing a comparable murder—or, at least, it does so as long as the morally relevant implications of each of these murders are comparable. But although common-sense morality includes such a constraint, no agent-neutral theory can accommodate one. The closest an agent-neutral theory can come to accommodating such a constraint is to give each agent the shared aim of minimizing murders and to give this aim absolute priority over their other shared aims, such as that of minimizing deaths. In that case, an agent would be prohibited from committing murder even to prevent thousands of deaths by natural causes. But such a theory would still permit an agent to commit a murder so as to prevent two others from each committing a comparable murder. Thus, the only way to accommodate an agent-centered constraint against murder is to give different agents different aims: for example, to give me the aim that I not commit any murders (or that I minimize the murders that I commit) and you the aim that you not commit any murders (or that you minimize the murders that you commit).

Now, if we understand consequentialism such that it can be agent-relative, then the following is a version of consequentialism that can accommodate constraints.

> *Agent-Relative Consequentialism (ARC)*: For any subject S and any available act φ: ($ARC_{P\text{-}Specifying}$) S's φ-ing is morally permissible if and only if there is no available alternative act ψ whose prospect she ought to prefer to that of her φ-ing. And ($ARC_{P\text{-}Making}$) the most fundamental permissibility-making feature of a permissible action is its lacking an alternative whose prospect she ought to prefer to that of its own.

I'll explain how ARC can accommodate agent-centered constraints in a moment, but first let me explain the notion of a prospect. As I noted earlier, the notion of φ's outcome assumes that there is just *one way* that the world *would* turn out if the agent were to φ rather than *several different ways* that it *could* turn out if she were to φ. This is problematic—see Hare's contribution to this volume (Chapter 18). So it's better to talk about the prospect of a subject's φ-ing. The prospect of a subject's φ-ing is the probability distribution consisting in the mutually exclusive and jointly exhaustive set of possible worlds that could be actualized by her φ-ing, with each possibility assigned a probability (an objective probability) such that the sum of these probabilities equals 1.[20] And, of course, if there is only one possible world that could be actualized by her φ-ing such that the probability that this world would be actualized if she were to φ is 1, then the prospect of her φ-ing is just the outcome of her φ-ing. Thus, strictly speaking, there's no need to talk (p. 12) about outcomes at all. We can just talk about the prospect of a subject's φ-ing, which in certain instances (where the probability that some possible world will result is 1) will be the outcome of her φ-ing.

Now, for the ARCist to accommodate a constraint against committing murder, she need only combine ARC with the view that, for any subject S, S ought, other things being

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

equal, to prefer the prospect of her refraining from committing murder to the prospect of her committing murder so as to prevent two others from each committing a comparable murder. The resulting view implies that Abbey is prohibited from murdering Abe even if this is the only way for her to prevent both Bertha from murdering Bert and Carla from murdering Carl. Also, it implies that Carla is prohibited from murdering Carl even if this is the only way for her to prevent both Abbey from murdering Abe and Bertha from murdering Bert. And it yields the same prohibition regardless of who's in the position of being the agent who can commit one murder so as to prevent the two others from each committing a comparable murder. But although this consequentialist view can accommodate constraints just as well as any nonconsequentialist view can, it will give a very different account of the fundamental right- and wrong-making features of actions and, thus, a different account of what makes it wrong to commit a murder so as to prevent two others from committing murder. To illustrate, consider the following nonconsequentialist view.

> *Kantianism (K)*: For any subject S and any available act φ: ($K_{P\text{-}Specifying}$) S's φ-ing is morally permissible if and only if S's φ-ing accords with Kant's categorical imperative. ($K_{P\text{-}Making}$) The most fundamental permissibility-making feature of a permissible action is its being in accordance with Kant's categorical imperative. And ($K_{CI\text{-}Specifying}$) S's φ-ing accords with Kant's categorical imperative if and only if, in φ-ing, S shows adequate respect for the rational, autonomous decision-making capacities of all those involved by treating them in only those ways that they could rationally consent to being treated and refraining from treating them in any way that they would reasonably not consent to being treated.[21]

Whereas the Kantian holds that what makes it wrong to commit murder even to prevent two others from each committing a comparable murder is that such an act fails to be in accordance with Kant's categorical imperative, the ARCist holds that what makes such an act wrong is that the agent ought to prefer the prospect of her refraining from performing it to the prospect of her performing it.

But although these two theories give different accounts of the fundamental right- and wrong-making features of actions, they can both hold that what ultimately grounds constraints is our duty to respect people and their capacity for rational, autonomous decision-making. Thus, as I see it, both theories can be deontological. For, as I see it, a theory is deontological if and only if it's a constraint-accepting theory that grounds its constraints in our duty to respect people and their capacity for rational, autonomous **(p. 13)** decision-making.[22] This is a duty, not to perform certain acts, but to have certain attitudes—for example, to regard persons as ends in themselves as opposed to "mere sites for the realization of value" (Lazar 2017, 582). Thus, the ARCist can be a deontologist by holding that the explanation for why a subject ought, other things being equal, to prefer the prospect of her refraining from committing a murder to the prospect of her committing that murder so as to prevent two others from each committing a comparable murder lies with the fact that she has a duty to respect each person's capacity for rational, autonomous choice—including that of her would-be murder victim's. And it's plausible to suppose that if she fulfills this duty, she will necessarily prefer, other things being equal,

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

the prospect of her refraining from murdering this would-be victim to the prospect of her murdering this person so as to prevent two others from each committing a comparable murder.

But although this view is deontological in the sense just defined, it is also a version of ARC in that it holds that the most fundamental permissibility-making feature of a permissible action is its lacking an alternative whose prospect is outranked by that of its own. By contrast, a theory such as Kantianism denies that what makes an act permissible has anything to do with whether its outcome or prospect is outranked by that of any alternative. And this difference between the two positions can best be illustrated by the following chart, where "→" stands for "grounds" and "**+**" stands for "conjoined with."

> Kantianism: The duty to respect persons → the duty to perform only those actions that accord with Kant's categorical imperative → the duty to refrain from committing a murder even to prevent two others from each committing a comparable murder.
>
> ARC_{Deon}: First, the duty to respect persons → the duty to prefer, other things being equal, the prospect of one's refraining from murder to the prospect of one's committing a murder to prevent two others from each committing a comparable murder. Second, the duty to perform only those actions whose prospects ought not be dispreferred to that of some alternative **+** the duty to prefer, other things being equal, the prospect of one's refraining from murder to the prospect of one's committing a murder to prevent two others from each committing a comparable murder → the duty to refrain from committing a murder even to prevent two others from each committing a comparable murder.

As this illustrates, the duty to respect persons is, on both views, what ultimately grounds our duty to refrain from committing a murder even to prevent two others from each committing a comparable murder. And to hold that there is a duty to refrain from committing a murder even to prevent two others from each committing a comparable murder is to accept a constraint against murder. Thus, both theories are constraint-accepting theories that ultimately ground their constraints in our duty to respect people and (p. 14) their capacities for rational, autonomous decision-making. And so both theories are deontological, as I've defined "deontology."

But, now, we may wonder: "Given that ARC_{Deon} and Kantianism are both deontological theories that ultimately ground their constraints in our duty to respect persons, is there any reason to prefer one to the other?" Indeed, I think that there are potentially three reasons for preferring ARC_{Deon} to Kantianism.

First, if we assume (admittedly controversially) that one can be morally required to do only what one has decisive reason to do (all things considered), then which theory we should prefer depends on whether we should accept the teleological conception of practical reasons. According to this conception of practical reasons, what makes an agent have more reason to φ than to ψ is that she ought to prefer the prospect of her φ-ing to the

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

prospect of her ψ-ing. Some, like myself, find this conception of practical reasons attractive as well as supportive of ARC.[23] We reason as follows. It is through our actions that we attempt to affect the way the world goes. Whenever we face a choice of what to do, we also face a choice of which of various possible worlds to attempt to actualize. Moreover, whenever we act intentionally, we act with the aim of making the world go a certain way. The aim needn't concern the causal consequences of the act in question, as the aim could be nothing more than to bring it about that one performs it. One could, for instance, intend to run merely for the sake of bringing it about that one runs. The fact remains, though, that for every intentional action there is some end at which the agent aims. It's natural, then, to think that the agent ought to act so as to make the world go as she ought to want it to go. And if there's no fact of the matter as to whether the world would go this or that way, then she ought, it seems, to perform the act whose prospect she ought to prefer to that of each of the available alternatives. This very simple and intuitive idea is the teleological conception of practical reasons. And if we find this conception attractive, it's natural to favor an agent-relative consequentialist conception of what we morally ought to do—one that holds that we ought to perform the option whose prospect we should prefer to that of all the available alternatives. At least, we should so long as we think that what we can be morally required to do is constrained by what we have decisive reason to do, all things considered.

Second, whether we should prefer ARC_{Deon} or Kantianism depends on whether we think that an agent's only ultimate concern should be for respecting people's autonomy. To illustrate, let's assume that the number *n* is sufficiently large that the world would be substantially better if Danielle were to murder Dan so as to prevent *n* deaths from natural causes than if she were to refrain from doing so and thereby allow these deaths to occur. But let's also suppose that Dan would reasonably refuse to consent to being murdered for the sake of preventing these *n* deaths given that none of those suffering any of these deaths by natural causes would suffer anything worse than what he would suffer in being murdered. Given these assumptions, Kantianism entails that Danielle is prohibited (p. 15) from murdering Dan. And although this would be acceptable if we thought that Danielle's only ultimate concern should be to refrain from violating anyone's autonomy, we clearly think that she should also have an ultimate concern for making the world better. And, thus, we think that the constraint against treating a person in a way that she would reasonably refuse to consent to has a threshold such that it will be permissible to treat her in this way if enough good is at stake. So, if we accept that agents should have, among other concerns, an ultimate concern for making the world better, then we should prefer ARC_{Deon} to Kantianism. For, unlike Kantianism, ARC_{Deon} implies that, if *n* is a sufficiently large number, it will be permissible for Danielle to violate Dan's autonomy for the sake of making the world better. Although the fact that Danielle would violate Dan's autonomy if she murders Dan and not if she refrains from doing so is a weighty reason for her to prefer the prospect of her refraining from murdering him, the fact that the world would be so much better if she were to murder Dan than if she were to refrain from doing so is an even weightier reason for her to prefer the prospect of her murdering him. And, thus, we

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

should find ARC_{Deon} more plausible than Kantianism if we think that the constraint against violating someone's autonomy has a threshold.

Third, the possibility of indeterminism seems to provide us with another reason for preferring ARC_{Deon} to Kantianism. To illustrate, imagine that Edith locks Ed up because she suspects that he'll commit murder if she doesn't. But assume that Ed has, as of yet, done nothing wrong. And assume that Ed has the sort of indeterministic freedom that implies that there is no fact of the matter as to what he would have done had Edith not locked him up. In other words, it's neither the case that (CF_1) Ed would have committed murder if Edith hadn't locked him up nor is it the case that (CF_2) Ed would not have committed murder if Edith hadn't locked him up.[24] So what's true about this possible murder is only both that the objective probability that Ed would have committed murder had Edith not locked him up was n ($0 < n < 1$) and that the objective probability that he would not have committed murder had Edith not locked him up was $1 - n$. Lastly, assume that Ed could have rationally consented to Edith's locking him up only if it had been a fact that he would had committed murder had she not locked him up—that is, only if n had equaled 1. Given these assumptions, Kantianism implies that Edith acted impermissibly in locking Ed up and that this is true regardless of how close n was to 1. But, some like myself, find it extremely implausible to think that Edith would have been prohibited from locking Ed up even if the objective probability that Ed would have committed murder had she not done so was, say, 0.99999.[25] So it seems that, again, ARC_{Deon} (p. 16) has the advantage. For, unlike Kantianism, it can account for the fact that, if n was close enough to 1, then Edith acted permissibly in locking him up even if she did, thereby, violate his autonomy given that, as we're supposing, Ed could have rationally consented to having been locked up only if n had equaled 1.[26]

What we've found, then, is that even if ARC_{Deon} and Kantianism accept all the same constraints and both agree that these constraints are ultimately grounded in our duty to respect people and their rational, autonomous decision-making capacities, the fact that the former is a version of ARC and the latter is nonconsequentialist can provide us with reasons for preferring the one to the other. But, of course, this was all based on several assumptions. So what if we reject these assumptions? And what if we think that Dreier is right about the difference between consequentialism and nonconsequentialism being merely notational? Will there still be good reasons to study the literature on consequentialism? I believe that there will.

First, many of the issues that have been most prominently discussed in the consequentialist literature are ones that are just as important for nonconsequentialists to address. This is because consequentialists and nonconsequentialists agree that the goodness of an act's consequences is something that matters; they just disagree on whether it's the only thing that matters. So all the issues concerning the goodness of an act's consequences that have been most prominently discussed in the consequentialist literature are just as much issues for nonconsequentialists. These include the following: (1) What exactly are the consequences of an action? In other words, what exactly does the relation between an act and a state of affairs need to be for that state of affairs to count as a consequence of that

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

act? See Dorsey's contribution to this volume (Chapter 5). (2) What's the relevant set of alternatives when it comes to assessing the consequences of our actions? For whether an act should be performed depends not only on its consequences, but also on how those consequences compare to those of the available alternatives. And that in turn depends on what counts as an available alternative. See Smith's contribution (Chapter 6). (3) To what extent can we compare various alternatives in terms of the goodness of their consequences? For instance, is there some number of headaches such that relieving all of them would be just as good as saving a life? See Norcross's contribution (Chapter 19). (4) If the consequences of our actions matter but we don't know what they are, how are we to assess our actions? See the contributions both by Bykvist (Chapter 16) and by Jackson (Chapter 17). (5) And what if it's not just that we don't know what the consequences would be, but that there's no fact of the matter as to what they would be? How, then, are we to assess our actions? See Hare's contribution (Chapter 18). (6) If what the consequences of our present actions would be depends on what we would freely choose to do in the future, how does this affect our assessment of our present actions? To illustrate, suppose that, as it happens, I would freely eat all the cookies in a box of cookies at t_2 if I were to open it at t_1. Should we, then, treat my eating all the cookies at t_2 as a consequence of my opening the box at t_1 and, thus, as a reason not to open it at t_1? And does this hold even if I could open it at t_1 and then freely choose to refrain from (p. 17) eating all the cookies at t_2? See the contribution by Cohen and Timmerman (Chapter 7). (7) What are the goods that we should seek to produce by our actions, and is there just one such good or many such goods? See de Lazari-Radek's contribution (Chapter 10). (8) Should we be concerned only with producing what's good for someone or also with producing what's good but good for no one? See Roberts's contribution (Chapter 25). And (9) how should we evaluate states of affairs, and do these evaluations depend on facts about the evaluator, her relationships, her physical location, the time at which she's making these evaluations, or the nature of the actual world that she inhabits? See Hammerton's contribution (Chapter 3).

Second, many of the issues that have historically divided consequentialists and nonconsequentialist are important for normative theorizing in general. For instance, consequentialists and nonconsequentialists have historically disagreed on whether there are supererogatory acts (see Archer's contribution; Chapter 14)), whether morality is very demanding (see Sobel's contribution; Chapter 11), whether morality allows us to be partial to those who are near and dear to us (see Jeske's contribution; Chapter 12), whether we're obligated to promote our own happiness whenever we can do so at no cost to others (see Cholbi's contribution; Chapter 13), and what sort of role, if any, should aggregation play in distributive justice (see Brink's contribution; Chapter 20). All these issues are important for normative theorists to address whether or not they are consequentialist.

Third, consequentialists have a long history of advocating social reform and have, as a result, given rise to several significant social movements. These movements are important in themselves but they cannot be adequately understood or assessed except in the con-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

text of the consequentialist views of their proponents. And so we must study consequentialism in order to properly understand and assess these movements.

Consequentialism's history of social reform dates back to the very first utilitarians—James Mill (1773–1836) and Jeremy Bentham (1748–1832)—and the first movement that they inspired: *philosophical radicalism*. The philosophical radicals were philosophically minded nineteenth-century British political radicals who believed that the principle of utility should be used to test all public policies and institutions. Consequently, their views were quite radical for their day. They advocated for prison reform, universal male suffrage, state-supported education for all, and reform of the British parliamentary system. Jeremy Bentham even argued in favor of legal protections for the interests of nonhuman animals, famously writing: "The question is not, Can they reason?, nor Can they talk? but, Can they suffer? Why should the law refuse its protection to any sensitive being?" (1789). And John Stuart Mill (James Mill's son) wrote in favor of women's suffrage. The utilitarians' propensity for pushing for social reform has continued to this day. For instance, Peter Singer is often credited with having started the modern-day animal advocacy movement. And the latest movement originating with utilitarians, viz., effective altruism, emerged over the last decade out of the ideas and writings of two young utilitarians at Oxford: Toby Ord and William MacAskill.

So, for these three reasons, it will be important for us to study consequentialism and its rivals even if, as it turns out, the difference between consequentialism and nonconsequentialism ends up being merely notational. With this is mind, I will now (p. 18) summarize what this volume has to offer in terms of helping us understand the field of consequentialism as it exists today as well as how points to new directions in which the field is likely headed.

3. A Very Brief Overview of This Volume

This volume is divided into four parts. Part I, entitled "Foundational Issues," contains nine chapters. Most of these deal with issues of concern to all normative theorists. For, as I've noted earlier, all normative theorists agree that consequences matter. And so they must all deal with issues about how we are to determine, compare, and evaluate our alternatives and their consequences. But a few of these nine chapters deal instead with issues that are particularly pressing for consequentialists. For instance, Paul Hurley's chapter is concerned with consequentializers: those who insist that all plausible accounts of what we ought to do can, and should, be put into a consequentialist form. He looks at the outcome-centered accounts of reasons, actions, and attitudes that they appeal to and argues that their putative plausibility trades on a conflation between two distinct senses in which we speak of actions as bringing about outcomes. In Chapter 8, Elinor Mason addresses the issue of the relationship between consequentialist principles and moral responsibility. She rejects the idea that we should turn to consequentialist principles to justify either our individual instances of blaming or our moral responsibility practices as a whole. Instead, she holds that our consequentialist principles must be constrained by our account of

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

moral responsibility. And, in Chapter 9, Christopher Woodard explains how consequentialists can provide plausible accounts of reasons for action. He argues that not only is the fact that an act would itself bring about a good outcome a reason to perform it, but also the fact that it would be part of a pattern of action that would bring about a good outcome can be a reason to perform it.

Part II, entitled “Objections,” contains eleven chapters. Each of these takes up a standard objection to consequentialism. Since I’ve already mentioned most of these in previous sections, I’ll just focus, here, on those that I have yet to mention. In Chapter 15, Alida Liberman considers the objection that consequentialists cannot adequately account for the moral force of promissory obligations. Krister Bykvist (Chapter 16) considers one of the main challenges facing act consequentialism: the ignorance challenge, the challenge that arises given that agents often lack both the empirical and evaluative knowledge needed to determine which of their actions would bring about the best consequences. Bykvist assesses the main responses to this challenge, but also argues that this challenge is as much a challenge for nonconsequentialists. In Chapter 21, Barry Maguire and Calvin C. Baker address the objection that consequentialism is alienating in that inhibits agents from fitting participation in the ideals of integrity, friendship, or community. They consider four consequentialist strategies for avoiding alienation objections and argue that none of them is fully satisfactory.

(p. 19) Part III, entitled “Forms and Limits,” contains six chapters. Each chapter concerns what form consequentialism should take. Hilary Greaves (Chapter 22) is concerned with global consequentialism, which holds that we should assess acts, rules, motives, decision procedures, and everything else open to deontic assessment directly in terms of their associated consequences, and she considers to what extent this form of consequentialism can deal with three standard objections to act consequentialism. Brad Hooker (Chapter 23) is concerned with rule consequentialism, which holds that we should assess sets of rules directly in terms of their associated consequences but assess acts in terms of whether or not they accord with the code of rules with best associated consequences, and he rebuts several old and new objections to the view. Now, rule consequentialism is a form of indirect consequentialism in that it assesses acts not directly in terms of their consequences but indirectly in terms of the set of rules that has the best associated consequences, but Julia Driver (Chapter 24) considers a different sort of indirect form of consequentialism: one that’s “indirect” in the sense that it allows that one can be a virtuous person by consequentialist standards even if one is not directly guided by consequentialist considerations. On such a view, a consequentialist must provide a consequentialist account of moral virtue, and Driver explores how this might be done. Melinda Roberts (Chapter 25) is concerned with whether consequentialism can take a form that credibly deals with cases in which we have to choose between bringing about outcomes involving populations of different sizes or populations of the same size but with a different set of individuals. Richard Yetter Chappell (Chapter 26) considers whether consequentialism should take a scalar, satisficing, or maximizing form, and argues that we should adopt all three, but only each with respect to a different deontic concept. That is, he argues that consequentialists should adopt a maximizing account of *the ought of most reason*, a satis-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ficing account of *obligation*, and a scalar account of *the weight of reasons*. Lastly, Joseph Mendola (Chapter 27) considers what form consequentialism must take if it's going to avoid self-defeat. He considers various possibilities, including consequentialist generalization, generalized act consequentialism, multiple-act consequentialism, cooperative consequentialism, and modally robust act consequentialism.

Part IV, entitled "Policy, Practice, and Social Reform," contains six chapters. These chapters all deal with consequentialism's role in, and relation to, several significant reform movements, and especially those calling for environmentalism, effective altruism, animal liberation, and women's liberation. For instance, Victor Kumar (Chapter 28) is concerned with effective altruism, and he argues that, despite its utilitarian origins, it is most plausible when divorced from utilitarianism. Judith Lichtenberg (Chapter 29) is also concerned with effective altruism. But she argues that this movement cannot be so easily separated from its consequentialist's origins, and she argues that effective altruism controversially presupposes that producing more good is always morally better than producing less. In Chapter 30, Tyler John and Jeff Sebo consider consequentialism's long and complicated history with the animal advocacy movement. On the one hand, consequentialists were some of the first moral theorists to recognize that nonhuman animals have moral status. But, on the other hand, many have questioned whether consequentialists (p. 20) can fully recognize the sort of moral status that each individual nonhuman animal has and the moral status that species themselves have. For they consider whether we should, perhaps, be concerned not only with the well-being of individual nonhuman animals but also with respecting their rights and with preventing the extinction of the species to which they belong. John and Sebo end up concluding that these seeming tensions between consequentialism and fully recognizing such moral status are overblown. In Chapter 31, Mark Buldofson and Dean Spears address various issues relating to public policy and consequentialism, especially issues that arise in connection with the environment. They suggest that the standard methods of economic policy analysis provide a good approximation for how consequentialists should evaluate various environmental policies. But they argue that this sort of analysis must be supplemented with methods for valuing animal well-being and making tradeoffs between this and human well-being. In Chapter 32, Samantha Brennan argues not only that there's an important historical connection between consequentialism and feminism but also that consequentialism, suitably revised and rebuilt, should count as a feminist approach to ethics. Lastly, in Chapter 33, Holly Lawford-Smith and William Tuckwell consider a challenge for act consequentialism concerning whether it's adequately equipped to deal with many of the public policy issues that we face today. Specifically, they address the no-difference challenge to act consequentialism. The challenge shows up in real-world cases, including those involving voting, global labor injustice, global poverty, and climate change. In all these cases, it seems that no individual's action makes a difference to the overall value of the outcome that we collectively bring about and yet we have the strong intuition that these actions are obligatory (or wrong). They work through a number of proffered solutions to this challenge, arguing that most fail but suggesting that two others may succeed.[27]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

References

Anscombe, G. E. M. 1958. “Modern Moral Philosophy.” *Philosophy* 33: 1–19.

Bentham, J. 1789. *An Introduction to the Principles of Morals and Legislation*. Oxford: Clarendon Press.

Brown, C. 2011. “Consequentialize This.” *Ethics* 121: 749–771.

Chappell, R. Y. 2015. “Value Receptacles.” *Noûs* 49: 322–332.

Dreier, J. 1993. “The Structures of Normative Theories.” *The Monist* 76: 22–40.

Dreier, J. 2011. “In Defense of Consequentializing.” In *Oxford Studies in Normative Ethics*, Vol. 1, edited by M. Timmons, 97–119. Oxford: Oxford University Press.

Estlund, D. 2017. “Prime Justice.” In *Political Utopias*, edited by K. Vallier and M. Weber, 35–35. Oxford: Oxford University Press.

Fraser, C. 2016. *The Philosophy of the Mozi: The First Consequentialists*. New York: Columbia University Press.

(p. 21) Hare, C. 2011. “Obligation and Regret When There Is No Fact of the Matter about What Would Have Happened If You Had Not Done What You Did.” *Noûs* 45: 190–206.

Hooker, B. 2000. *Ideal Code, Real World*. Oxford: Oxford University Press.

Howard-Snyder, F. 1993. “Rule Consequentialism Is a Rubber Duck.” *American Philosophical Quarterly* 30: 271–278.

Lazar, S. 2017. “Deontological Decision Theory and Agent-Centered Options.” *Ethics* 127: 579–609.

Lewis, D. 1973. *Counterfactuals*. Oxford: Blackwell.

Norcross, A. Forthcoming. *Morality by Degrees: Reasons without Demands*. Oxford: Oxford University Press.

Portmore, D. W. 2011. *Commonsense Consequentialism: Wherein Morality Meets Rationality*. New York: Oxford University Press.

Portmore, D. W. 2019. *Opting for the Best: Oughts and Options.* New York: Oxford University Press.

Regan, D. 1980. *Utilitarianism and Co-operation*. New York: Oxford University Press.

Scheffler, S. 1985. “Agent-Centred Restrictions, Rationality, and the Virtues.” *Mind* 94: 409–419.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Sen, A. 2000. “Consequential Evaluation and Practical Reason.” *The Journal of Philosophy* 97: 477–502.

Sinnott-Armstrong, W. (2015). “Consequentialism.” In *Stanford Encyclopedia of Philosophy* (Summer 2015), edited by Edward N. Zalta. **http://plato.stanford.edu/archives/sum2015/entries/consequentialism/**.

Stalnaker, R. 1984. *Inquiry*. Cambridge, MA: MIT Press.

Vessel, J.-P. 2003. “Counterfactuals for Consequentialists.” *Philosophical Studies* 112: 103–125.

Notes:

([1]) The relevant outcomes needn’t be those of the acts themselves. After all, rule consequentialists hold that the deontic statuses of our actions are a function of the ranking of the outcomes resulting from our accepting (or complying with) various sets of rules that either permit or prohibit these actions.

([2]) A subject’s φ-ing maximizes the good if and only if there is no available alternative act ψ whose outcome is better than that of her φ-ing.

([3]) Contrast this with *Theological Consequentialist Voluntarism (TCV)*: For any subject S and any available act φ: ($TCV_{P\text{-}Specifying}$) S’s φ-ing is morally permissible if and only if God doesn’t forbid her from φ-ing. ($TCV_{P\text{-}Making}$) The most fundamental permissibility-making feature of a permissible act is its not being forbidden by God. And ($TCV_{F\text{-}Specifying}$) S’s φ-ing is forbidden by God if and only if there is an available alternative act ψ whose outcome is better than that of her φ-ing. Like TAC, TCV holds that a permissibility-making feature of a permissible action is its lacking an alternative whose outcome is better than its own. But, unlike TAC, TCV denies that this is *the most* fundamental permissibility-making feature of a permissible action. For, on TCV, this is a permissibility-making feature only in virtue of the fact that God forbids all and only those acts that fail to maximize the good.

([4]) This is a contemporary development. When the term “consequentialism” was originally coined by G. E. M. Anscombe (1958), the thought was that the consequentialist cares about only the causal consequences of actions.

([5]) The assumption that there’s just *one way the world would be* if S were to φ rather than just *several different ways that it could be* if S were to φ is known as *counterfactual determinism*. It is, admittedly, quite controversial—see Hare, this volume, Chapter 18. And if this assumption is false, we’ll need to replace “outcome” with “prospect” in the earlier formulation of TAC. I’ll explain the notion of a prospect in section 2.

([6]) See Paul Hurley’s contribution to this volume (Chapter 2) for a discussion of Dreier’s claims.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([7]) Not all agree. Frances Howard-Snyder (1993) has argued that rule consequentialism is a rubber duck in that it is no more a species of consequentialism than a rubber duck is a species of duck. Also, as we'll see later, some may question whether RC is consequentialist given that it must eschew a commitment to making the world as good as possible (see, e.g., Portmore 2019, 41–44). For more on rule consequentialism, see Hooker's contribution to this volume.

([8]) According to Hooker (2000), if two or more codes are tied for first place in terms of their expected goodness, the ideal code is the one closest to our actual moral code.

([9]) For more on scalar consequentialism, see Alastair Norcross's forthcoming book *Morality by Degrees: Reasons without Demands*.

([10]) The hedonic utility of an act equals the total amount of hedons it produces minus the total amount of dolors it produces, where a hedon is a unit for measuring the intensity and duration of an episode of pleasure produced by an act and where a dolor is a unit for measuring the intensity and duration of an episode of pain produced by an act. The greater the intensity or duration of an episode of pleasure (or pain), the greater the number of hedons (or dolors) that episode contains. And these units are such that any act that produces exactly as many dolors as hedons has a hedonic utility of zero.

([11]) The example is just a more concrete version of Donald Regan's case of Whiff and Poof—see his 1980.

([12]) Admittedly, classical utilitarianism does require each agent to produce the best world that she can produce given what others are doing. But it seems that what ultimately matters is that the world be as good as possible (and, thus, that Slice and Patch together produce O_1), not that each produces the best world that she can produce given what others are doing (and, thus, not that they together produce either O_1 or O_4). Indeed, it seems that the only reason that we want each agent to produce the best world that she can produce given what others are doing is that we want the world to be as good as possible and see this as a potential means to that end. Thus, to care about each agent's acting so as to produce the best world that she can produce given what others are doing and to care about this regardless of whether they do so in a way that brings about the best world that they can together produce (that is, regardless of whether they together produce O_1 or O_4) involves the mistake of valuing something that is only instrumentally valuable as if it were noninstrumentally valuable.

([13]) See, for instance, Regan (1980) and Portmore (2019). And see Mendola (Chapter 27, this volume) for further discussion of this issue.

([14]) The following discussion is inspired by Chappell (2015).

([15]) I'm not concerned with whether this is plausible. I'm just helping the reader get a feel for the theory's implications.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

([16]) What ultimately matters to us (or to some individual) are all and only those things that we (or that individual) should nonderivatively care about—that is, all and only those things that we (or that individual) should care about independent of its relation to anything else that we (or that individual) should care about.

([17]) It has also played an important, though perhaps less prominent, role in other traditions. See, for instance, Amartya Sen's discussion of the Hindu text the *Bhagavad Gītā* in his 2000 and Chris Fraser's discussion of the Chinese text the *Mozi* in his 2016.

([18]) Note that indirect consequentialist theories such as rule consequentialism are agent-relative theories even though they appeal to an agent-neutral ranking of outcomes. That is, they give different agents different aims. For instance, rule consequentialism gives me that aim that I not violate the ideal code, but you the aim that you not violate the ideal code.

([19]) For more on such fault lines, see Nair's contribution to this volume (Chapter 4).

([20]) The objective probability that some event will (or would) occur is the percentage of the time that it will (or would) occur under identical causal circumstances—circumstances where the causal laws and histories are exactly the same. Of course, if we want to give an account of permissibility in the evidence-relative sense rather than the fact-relative sense, we need only replace these objective probabilities with evidential probabilities.

([21]) I'm not interested in whether this is the correct interpretation of Kant. That's why I call this *Kantianism* rather than *Kant's View*.

([22]) I owe both the idea that we should define deontology along these lines and the idea that a deontologist, so defined, could be a proponent of ARC to Jake Zuehl, who suggested them to me via email in May 2017.

([23]) In my 2011, I argue that we should accept both the teleological conception of practical reasons and the view that one can be morally required to do only what one has decisive reason to do (all things considered). And I argue that ARC follows from these two views.

([24]) On the assumption that Ed's behavior is not causally determined, there seems to be nothing about the world that makes either these two counterfactuals true. Thus, I'm assuming that we should accept something like the Lewis-Stalnaker semantics for counterfactuals, where neither CF_1 nor CF_2 is true. Either they are both false (as they are on Lewis's theory) or they both have indeterminate truth values (as they do on Stalnaker's theory). See Lewis (1973) and Stalnaker (1984). Here, I'm relying on interpretations of Lewis and Stalnaker given in Hare (2011) and Vessel (2003).

([25]) Given the indeterminacy involved, it is not true that Ed would have committed murder if Edith hadn't locked him up. Nonetheless, it is, we'll suppose, true that the objective

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

probability that he would have committed murder if she hadn’t locked him up was 0.99999.

([26]) For more on this, see chapter 7 of my 2019.

([27]) For helpful comments on earlier drafts, I thank Josh Glasgow, Frank Jackson, Barry Maguire, Joseph Mendola, Holly Smith, and Christopher Woodard.

Douglas W. Portmore

Douglas W. Portmore is Professor of Philosophy at Arizona State University. His research focuses mainly on morality, rationality, and the interconnections between the two, but he has also written on blame, well-being, moral worth, posthumous harm, moral responsibility, and the nonidentity problem. He is the author of two books: Commonsense Consequentialism: Wherein Morality Meets Rationality (Oxford University Press, 2011) and Opting for the Best: Oughts and Options (Oxford University Press, 2019).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Index

The Oxford Handbook of Consequentialism
Edited by Douglas W. Portmore

Print Publication Date: Dec 2020 Subject: Philosophy Online Publication Date: Oct 2020

(p. 655) # Index

ability 81, 148n.20, 150, 202, 238, 243, 247n.27, 248, 258, 279, 286, 293–294, 339, 361, 363, 396, 476n.5, 489n.38, 514, 525–526, 539, 544, 550, 556, 569, 602, 623, 625
joint 526
physical 126
rationality-ability principle 155–156
time and 136, 146, 146n.13
to do otherwise 75, 140n.6
ableist 627
act:
descriptions 103, 106–107, 123, 465
entailment 74, 87, 89–90, 108, 114, 119, 119nn.20–21, 134, 135–136n.42, 143, 143n.11, 144
generation 118–130, 118n.18, 126n.30, 133–135, 136n.42
guidance 153, 158, 171, 171n.14, 197, 200, 202, 275, 310–317, 316n.4, 320, 331–343, 369, 503–504, 511, 628
individuation 103, 115n.8, 118, 118n.15, 122
sequences 115, 127n.33, 128, 136
tokens 118, 118n.15, 118n.17, 120–121, 120n.22, 121nn.24–25, 124–137, 126n.32, 132nn. 35–36, 134nn.38–39, 135n.40, 136n.42, 183, 187, 190, 192, 430
trees 120–136, 120n.22, 128n.34, 134n.38
types 116, 118n.16, 121, 121n.25, 124–133, 427, 519
versions 76n.13, 113–117, 119n.20, 128, 133, 135–136, 135n.32
actions:
act consequentialism. *See* consequentialism, act
component 128–133, 136
compound 336–340, 434
co-temporal 114–115, 127n.33, 128, 136
group 517, 520–527, 522n.23, 644
maximal and nonmaximal 135n.42, 136n.42, 143–146, 143n.11, 148n.20, 150, 150n.23, 152–153, 156, 158, 514, 572n.26
standard story of 37, 37n.26
act/omission distinction. *See* doing and allowing
actualism 74n.10, 81–82, 84n.19, 87–91, 115, 139–159, 188n.15, 434, 434n.21, 435, 435n.22, 565, 572n.26, 581n.50 (*see also* moral, actualism)

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

bad behavior objection to 151, 157–158
nonratifiability problem and 152, 157–158
not demanding enough objection to 151
Adams, C. J. 573, 573n.30
Adams, R. M. 257n.8, 409, 426, 431–432, 432n.15, 468–469, 471
Adler, M. 594, 601, 603
admirability 250, 541, 556, 560–562
aesthetics 93, 95, 402, 415, 543, 596, 598
Agamemnon 368
agent-centered. *See* agent-relative
agent-centered constraints. *See* constraints
agent-centered options. *See* options, agent-centered
agent-centered prerogative. *See* options, agent-centered
agent-centered restrictions. *See* constraints
agent-neutral:
commitment 25–26, 28–29, 34
consequentialism and 47, 50, 62, 301–302, 303n.24, 342, 379, 382, 397, 624
consequentializing. *See* consequentializing, agent-neutral
considerations 68n.4
facts 408
interpretation of deontic constraints and special duties 49, 50n.5, 51
(p. 656) point of view 25, 27n.5, 28–29
ranking of outcomes 5, 6, 10n.18, 27, 28n.16, 43, 55
reasons 249–250, 265, 381
rules 48–49, 48n.4, 55, 63
theories 10–11, 48, 50–51, 63, 383n.5, 443–444
value 29, 36, 52–54, 61, 62, 64, 179, 193, 247, 302, 381, 442, 624
agent-relative:
act consequentialism and 11, 14, 25, 25n.1, 29, 29n.8, 34–36, 38, 41, 47, 50, 51n.8, 52, 54, 59n.20, 60–63, 63nn.24–25, 302–303, 303n.24
commonsense morality and 2, 28–29, 49–50
consequentializing. *See* consequentializing, agent-relative
constraints. *See* constraints
goodness. *See* agent-relative, value
permissions. *See* permissions
ranking of outcomes 5, 28, 28n.6, 35, 46, 46n.1, 50–51, 54–55, 416
reasons 249–250, 264–265, 381
requirements 63, 304n.28
rule consequentialism and 10n.18
rules 48–50, 54–55, 57–58
teleology 60
theories 10, 48–50, 48n.4, 50n.5, 54, 59, 63–64, 302, 383n.5
values 47n.1, 48, 50n.6, 51–54, 60–62, 179, 192–193, 301, 303, 443–444, 470–471, 625
agglomeration 73–78, 82–91, 433n.19, 435n.22
aggregation:
axiological. *See* aggregation, value
contractualism and 361, 361n.7, 385–386
cost 598

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

deontic 363, 364, 364n.9
distal 379, 391–394, 398
distributive justice and 17, 379–388
harms and 364–366, 368n.11, 390, 459
interests and 378
interpersonal 359–360, 363, 376, 378–379, 382–391, 391n.7, 398
intrapersonal 376, 387, 392, 392n.8
limited 361, 361n.7
local 379, 391–396, 398
prudential 383
selective 379, 390, 394–395, 398
unrestricted 364, 379, 390–398, 391n.7
value 353, 363–376, 442, 477, 592, 595, 600–601, 611
Alexander, L. 383n.3
alienation 18, 174n.17, 175, 225n.7, 244, 401–416, 464, 468, 471, 540
Almassi, B. 643
Almeida, M. 147n.17
Alston, W. P. 211n.46
alternatives:
act trees as 123, 127–128
bottom-most act tokens as 122–124
consequentialism and 113, 115
constraints on what counts as one 116
highest normatively significant act tokens as alternatives 122, 124–127
multiple versions of 16, 18, 113–115, 117, 117n.14, 118n.15, 515
altruistic actions. *See* actions, altruistic
altruism 17, 19, 146, 253–258, 381–382, 385, 431, 502, 507–508, 531–546, 548–563, 566–567, 577, 579, 582–583, 610, 641 (*see also* effective altruism)
Alvarez, M. 180n.4
Anderson, E. 37n.27, 539
Andreou, C. 643
animal:
agriculture 565, 567, 569, 571–572, 576–578, 581
extermination 207, 581–587
pleasure 214
rights 564, 571, 575–576, 579, 617
animals:
attitudes and. *See* attitudes, animals and
domesticated 564–567, 569, 570–571, 574–578, 582, 584, 587, 599
morality and 17, 19, 20, 253, 261, 322–326, 335, 359n.5, 376, 468, 476n.6, 531, 538, 542, 564–588, 592–593, 595–596, 598–599, 605–611, 618 (*see also* moral, status)
wild 565–567, 569, 579–581, 583–587
Annas, J. 470n.6
Anscombe, G. E. M. 2n.4, 36, 42–43, 118, 470, 501n.3, 554n.24
(p. 657) Appleby, M. 607
Åqvist, L. 115n.7, 515n.5
arbitrariness 114, 130, 392–395, 397, 427, 429n.10, 433n.18, 434n.21, 471, 501, 505–506, 509, 534, 541, 559, 644

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Archer, A. 17, 115n.7, 269, 272, 280, 280nn.13–14, 281n.15, 282, 448
Aristotle 31, 202n.18, 214n.58, 321n.6, 465
Arneson, R. 166, 389, 398n.13
Arnold, M. 212, 212n.50, 212n.53
Arntzenius, F. 639
Arpaly, N. 403, 468, 468n.5, 470
Arrhenius, G. 474n.2, 492n.45, 570n.22
Arrow, K. 599
Ashford, E. 463n.1
asymmetry:
  between action and inaction 500
  between appropriately joining and defecting from group acts 522–525
  moral 388–391, 398, 525
  procreative 481, 570
  self-other 259, 267
attitudes:
  agential 411
  animals and 531, 571–576, 578, 581–582, 587
  blame and 163–164, 455n.10, 503n.7
  fittingness and 503n.6, 506, 508
  irrational 503
  justice and 469
  outcome-centered accounts of 18, 35, 37–39, 41, 43
  partial 241, 244–246, 618, 625
  praise and 164, 226
  propositional 35–37, 43, 503n.6
  reactive 164, 226, 455n.10, 502–504, 506
  requirements with respect to 13
  value and 203, 205–206, 208–212, 241, 469
  vicious 571
Augustine 550
Austin, J. 425
autonomy 14–16, 116n.9, 262, 263, 266, 304n.28, 411, 499, 557, 581–584, 624, 627–628
average principle 477–482, 488–489
axiology 46–47, 50–51, 51n.8, 54, 56, 59–60, 62–64, 63n.26, 106n.19, 140, 198, 203, 259, 278, 358, 363–364, 366, 368n.11, 376, 402, 409n.14, 423, 427–428, 430n.12, 438–439, 570, 580, 592–593, 595, 601, 604, 608
Aydede, M. 212n.48
Babbitt, S. 628
Baber, H. E. 628–629
Bacharach, M. 188n.15
Bader, R. 407, 479n.10
Badhwar, K. N. 245, 245nn.21–22, 463
Baier, A. 618
Baier, K. 452n.8
Baker, C. 18, 174n.17, 401
Baker, D. 151n.27
balancing:

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

interpersonal 378, 383–384, 383n.5, 386–388, 391, 397
intrapersonal 383–384, 387, 397
Bales, R. E. 426n.5, 452
Banzhaf, H. S. 602
Bardsley, N. 187
Barnes, Elizabeth 395n.10, 531
Barnes, Eric 308n.30
Barnett, Z. 641
Baron, M. 245n.20
Barron, A. 607
Barry, B. 554, 554n.24
Barry, C. 641, 643, 645, 648–650
Bartky, S. L. 627
Bastian, B. 573–574, 573nn.31–32, 574n.33
Beckstead, N. 577n.44
Beerbohm, E. 640
beneficence 107, 189n.16, 256, 262, 279, 327, 448–500, 522, 541, 542, 546, 646
Benn, C. 121n.25, 285, 286
Bennett, J. 98n.10, 102, 102n.16, 103n.17, 116n.11, 163n.2
Bentham, J. 17, 27, 199, 290n.2, 380n.2, 425, 531, 557, 564, 566, 566n.6, 617nn.2–4
Bergström, L. 98n.8, 113, 113n.1, 114n.2, 115, 116n.12, 118n.16, 514n.2
Berkey, B. 610
Bernstein, S. 100, 100n.13, 101n.14, 637–638, 641
Berridge, K. 212, 212nn.54–55, 213n.56
bias 214, 298, 424, 447, 538, 559, 568, 578, 599
Bilz, K. 576n.42, 576n.43
(p. 658) blame 18, 61, 162–177, 226–227, 234n.16, 293, 306, 313–314, 320–321, 454–456, 455n.10, 466, 469, 500, 502, 506–507, 510, 555n.28
blameless wrongdoing 168–169, 168n.9, 169n.10, 170n.12, 250, 316, 320, 323–324, 578, 586
blameworthiness 119n.19, 121, 121n.24, 131, 133, 151n.28, 162, 167–168, 170–172, 180–181, 181n.6, 226, 285, 314, 316, 320–321, 324, 333, 335, 500, 502–503, 503n.7, 506n.9, 507–510, 513, 553, 555n.28
Blum, L. 406
Bradley, B. 175n.18, 190n.18, 275n.6, 467, 505
Braham, M. 639
Bramble, B. 211n.45, 211n.47
Brandt, R. 98n.9, 163n.2, 165, 166n.6, 189, 211n.46, 303n.25, 304n.26, 449n.5, 568n.10
Bratanova, B. 574n.33
Bratman, M. E. 521n.21
Brennan, G. 406, 411, 425, 640
Brennan, J. 639–640
Brennan, S. 20, 616, 622, 630–631
bringing about an outcome 18–19, 28, 36n.23, 37–40, 42–43, 63, 147, 173, 267, 318–319, 324, 452, 506, 537, 558, 616n.1
Brink, D. O. 17, 243–245, 245n.19, 247, 249n.33, 360, 378, 382, 389, 392n.9, 398n.13, 425, 540
Broad, C. D. 381
Brook, R. 55n.14

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Broome, J. 33n.17, 51n.8, 52n.11, 55, 60, 477n.8, 480, 480n.15, 486, 488n.34, 489n.36, 492n.45, 570n.21, 593, 642–643
Brown, C. 3, 30n.13, 47n.2, 67n.2, 70–71, 71n.6, 76n.13, 114nn.2–4, 115nn.6–7, 116n.9, 119nn.20–21, 121n.26, 133, 143n.11, 144, 150n.22, 156n.32, 433n.19
Browning, H. 607
Brudney, D. 401n.2
Brundtland, G. 602
Buchak, L. 593n.1
Budolfson, M. 564n.1, 570n.21, 592–593, 593n.1, 598, 601, 608, 610–611, 641
Bunzl, M. 637
Burgess-Jackson, K. 618n.7, 620
Buttlar, B. 574n.35
Bykvist, K. 16, 18, 114nn.2–4, 115n.7, 118n.15, 119n.19, 144n.12, 150n.24, 310, 317–318, 320, 323–324, 326–327, 444
Byron, M. 276n.7
Calder, T. 467
Calhoun, C. 630
Card, C. 624
Cariani, F. 84n.18, 85, 148n.19
Carlson, E. 113, 114nn.3–4, 115nn.6–7, 119n.19, 144n.12, 344n.1, 364–365
Carlyle, T. 214, 214n.58
Castañeda, H.-N. 113, 115, 141–143, 514n.2
categorical imperative 2, 3, 6, 12–13, 116, 376
causation 101, 350–351, 353, 635–636, 638–639, 646
  backward 97
  counterfactual theory of 100, 636–638, 647
  individual 643
  influence theory of 636–638
  joint 643
  productive theory of 636
Chan, K. 571, 574n.34, 605n.4
Chang, R. 201, 201nn.15–16, 201n.17, 375
Chappell, R. Y. 8n.14, 19, 184, 239n.6, 407, 428n.9, 438n.25, 439, 452n.9, 498, 503, 505–507
character 31–33, 40, 93n.1, 151, 165, 179, 190, 293, 328, 335, 380, 411, 413, 425, 428, 430–431, 431n.13, 438, 463–471, 513, 515, 525, 554, 555n.28, 560, 568, 571, 581, 619–620, 630, 635
Cholbi, M. 253, 296n.12, 298–299, 299nn.16–17
climate 20, 428, 523, 525–526, 539, 542, 556, 570n.21, 594–596, 602, 605–606, 609–611, 635, 639, 641–643, 646, 647–648, 650
Cocking, D. 173n.15, 246–247, 246n.24, 247n.25, 412–413, 568n.16
Cohen, G. A. 406, 415
Cohen, Y. 17, 74n.10, 114n.4, 115, 133, 139, 140n.6, 143, 147n.17, 148n.19, 149, 151n.29, 158, 340, 444, 572n.26
Cohon, R. 470n.6
collective action 459, 593, 640–641, 649–650
(p. 659) commensurability 201, 359, 367–376
commonsense morality 10–11, 28–29, 37, 47, 49–50, 59, 64, 73, 194, 269, 273, 276, 292, 297, 299n.17, 303, 306, 381, 427n.6, 430, 437, 447n.4
comparability 201–202, 234, 358–376, 379, 398, 478n.10, 480n.14, 481n.20

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

compensation 378–379, 383, 383n.5, 384–385, 387–388, 397–398
Connelly, M. 595
conscientiousness 311, 318, 319–321, 324–325, 326n.9, 329, 437, 468, 510, 572, 574, 575, 579
consequence argument against compatibilism 104
consequences:
act itself and 2, 80, 98, 101–103, 105–106
actual 94, 239, 444–445, 453–454, 466
causal 2, 2n.4, 14, 79, 98n.10, 114, 122–123, 133, 185, 345
constitutive 39n.29, 85, 241, 245, 300
expected 239, 441
moralized approach to 95–96
probable 239
traditional approach to 97–101
consequentialism:
act:
agent-relativity and 5, 6, 10–16, 13n.22, 14n.23, 25n.1, 29, 29n.8, 34, 35–36, 38, 41, 46–47n.1, 47–48, 51n.8, 52, 54, 59n.20, 60–61, 63, 63nn.24–25, 302
alienation objection and 18, 174n.17, 175, 225n.7, 244, 401–416, 464, 468, 471
compulsory self-benefit objection and 253–268
constraints and. *See* constraints, act consequentialism and
demandingness objection and 17, 169, 221–237, 240, 253, 261–264, 272–273, 276–277, 280, 285–286, 293n.8, 299, 343, 386–387, 437, 452, 498–500, 505, 510, 517, 549–553, 551n.11, 587, 593, 617, 626, 629
dual-ranking 256, 264–267, 270, 276–280, 502
generalized 19, 520–521
harmonious 403–406, 409
ignorance challenge and 18, 310–329, 447–448, 513, 587
incorrect verdicts objection and 47, 50, 59, 67, 71, 423, 430, 436, 438, 507
integrity challenge and 18, 67, 173–175, 412, 463
modally robust 19, 524–525, 525n.31
no-difference challenge and 20, 101, 103, 634–650
position-relative 46–47, 46n.1, 60, 62
promises and. *See* promise, act consequentialism and
reasons and 180, 182–187, 189, 191
self-defeatingness objection and 19, 423, 425, 436, 438–439, 514–519, 523, 526 (*See also* self-defeat)
separateness of persons objection and 360, 378–398, 540
silence objection and 424–425, 428, 436, 438
standard 68, 74, 254–255, 525
straightforward 289–291
traditional 1–3, 5, 310, 416, 506, 519, 523–524
treating people as mere receptacles for pleasure and 8, 564
bullet-biting and 155, 173–174, 242, 255, 272, 311, 313, 323–324, 329, 356, 635, 644
compellingness of 33, 60–61, 69–70, 72, 78, 259, 295, 427n.6
consequentializing argument for 26, 29–34, 38
cooperative 19, 524–525, 525n.21
definition of 1, 4–10, 179, 180n.1, 182–184, 194, 239, 240n.8, 270, 380, 553–554, 559, 616n.1, 617

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

direct 246, 283, 409, 427
distribution-sensitive 386, 442
global 19, 403, 409–410, 410n.16, 411, 415–416, 423–439
indirect 10n.18, 19, 166, 185–190, 192, 243–248, 250, 270, 282–286, 403n.6, 520, 561, 568
leveled 403, 409–415
maximalist 26n.2
maximizing. *See* maximizing
(p. 660) motive 189, 190, 190n.18, 256–258, 561
multiple-act 19, 522–523, 526–527
objective 94–95, 111, 171–172, 243, 244, 244n.16, 245, 314, 316, 465–466, 469
rule 1n.1, 4, 4n.7, 6, 9, 10n.18, 19, 84, 180n.1, 185–190, 192–193, 239, 239n.5, 239n.7, 283, 289, 290n.1, 298, 303–307, 423, 426, 427n.6, 441, 452, 454–460, 457nn.11–12, 471, 520, 521n.20, 555n.28, 561, 561n.50
collapse objection and 9, 188, 305, 561, 561n.50
conflicting rules and 187
incoherence objection and 9, 180n.1, 305, 426, 457–458
overarching commitment to maximizing the good and 9, 457, 457n.11
promises and. *See* promises, rule consequentialism and
reasons and 187–190, 459
satisficing 84, 239, 258–259, 270, 274–276, 276n.8, 278, 280, 430, 452n.9, 498, 504–507, 509–510
scalar 4, 4n.9, 273, 293n.8, 498, 501–504, 508–509, 553, 553n.20, 558, 565n.4
sophisticated 169n.10, 244–247, 250, 261n.12, 465, 567–569
subjective 171, 171n.13, 243–246, 244n.16, 466
two-level 166, 568, 626
consequentialist generalization 19, 519–520
consequentializing 18, 25–43, 54, 57, 85–86, 543
agent-neutral 27–30
agent-relative 29, 34, 36, 38
consequentialist argument for 34–43
conservation 490–491, 492n.47, 565–566, 569, 580, 585–586
constraints 10–16, 28–29, 43, 47, 49, 50–51, 54, 55n.15, 58–59, 59n.20, 61, 68–73, 68n.4, 79, 145, 190–194, 190n.19, 209, 381–382, 382n.3, 384, 505–507, 568–569, 611, 630–631, 635
contractualism 9, 233, 296, 296n.12, 298n.15, 302, 361–362, 378–379, 384–386, 389, 398, 459, 522, 566
cooperation 19, 65, 73, 73n.8, 225, 257, 290n.3, 301, 383, 447, 513–527, 545, 584
Copp, D. 226n.8, 459–460
costs:
active compliance 232–233
internalization 283, 283n.16, 450
passive compliance 233
psychological 424, 450
self-imposed 500
Cottingham, J. 239, 239n.3, 560n.46
counterfactual determinism 2n.5, 312, 434n.21
counterfactuals 24, 80, 88n.23, 100, 135, 140, 140n.7, 149, 149n.21, 312, 354
Cowen, T. 570n.20, 599, 610
credences 311, 320–321, 325, 326n.10, 329, 332–333, 336, 338

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Crisp, R. 179, 180n.3, 214n.59, 273n.5, 439, 442, 464
criterion of rightness 303, 340, 358n.3, 452, 454–458, 504, 508, 561n.51, 565, 567–568, 572, 580
Crompton, L. 617n.2
Cullity, G. 52n.12, 61, 641, 646
Cummiskey, D. 27n.5
Darwall, S. 68n.4, 204, 204n.28, 234n.16, 301n.20, 405
Dasgupta, P. 485n.13, 605n.4
Davidson, D. 118
Davis, A. Y. 576, 576nn.39–41
Dawkins, M. 607
Dea, S. 618n.6
Deaton, A. 557n.35, 558, 558n.40, 595, 601
De Brigard, F. 101n.14, 214n.63
decision procedures 19, 157, 171, 225, 225n.3, 243, 275, 293, 303n.24, 411, 425–427, 426n.4, 429, 431–432, 432n.14, 436–438, 445–458, 463, 507, 513, 525n.31, 561n.51, 565, 567–569, 572, 574, 577–578, 580, 583, 586–587, 626
decision theory 84, 313n.2, 314n.3, 317, 333–334, 408, 427, 434n.21, 592–593, 604, 621
deliberation 31, 57, 180–181, 211, 243–245, 247, 250–251, 264, 267, 299n.16, 304, 464–465, 510, 627
Delon, N. 567n.14, 580n.49
(p. 661) demandingness objection. *See* consequentialism, act, demandingness objection and
deontically equivalent 30–31, 38
deontic:
- actualism. *See* actualism
- aggregation. *See* aggregation, deontic
- averagism 83–84
- concepts 19, 274, 501, 501n.3, 508–510
- constraints. *See* constraints
- equivalence thesis 30, 38
- evaluation 25–27, 26n.3, 29–30, 35–36, 38, 43, 409n.14, 423, 427, 428–431, 436, 438–439, 464, 470
- logic 84–85, 87, 114, 133, 140n.4, 141, 144, 147, 159, 188n.15
- maximin possibilism. *See* possibilism
- monism 508–509
- options. *See* options, agent-centered
- pluralism 498, 508–511
- principles 115, 134, 403n.7
- properties 121n.24, 402
- statuses 1, 1n.1, 4, 6, 26, 28, 32–34, 39, 47, 52, 57–58, 68, 80–84, 86, 139, 144, 266–268, 284, 285, 456, 501–502, 501n.4
- values 123, 140n.5, 317
- verdicts 2, 31, 50, 54, 59, 507

deontologizing 31–34, 31n.16
deontology 12–14, 13n.22, 26, 32, 68, 113–114, 116–117, 122–123, 125–126, 135n.39, 198, 240, 246n.23, 291, 300n.19, 303–304, 303n.24, 307, 333, 380–382, 382n.3, 407, 555, 560, 562, 566, 581, 583, 626, 630–631
desert 167, 183, 231, 234, 303, 313–314, 320, 327, 538, 545, 573, 585, 622
desire-based theories. *See* desire satisfactionism

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

desire satisfactionism 197, 208–210, 209n.41, 212n.9, 213, 380, 565, 603–604, 627–628
determinism 104, 163–164 (*see also* counterfactual determinism)
Diamond, C. 573, 573n.28
Dickens, M. 567n.11
Dietrich, F. 67n.2
Dietz, A. 188
difference making 58, 97–98, 250, 535, 541, 549n.4, 556, 565, 635–648
difference principle 385, 387–389, 389n.6
dilemmas 57, 63, 73n.8, 157, 368–369, 432n.17, 470, 484, 516, 538, 551, 559n.42, 624
disabilities 390, 531, 619
disaster prevention 175, 189, 291, 315, 436, 450–451, 453, 457–458, 520n.17
dispositions 150, 244–247, 380, 411, 437, 447–448, 450, 463, 465, 471, 541, 625
distributive justice 17, 327, 378–380, 383, 386–397, 601, 611
doctrine of double effect 172–174, 173n.15, 328
doing and allowing 175–176, 507, 566 (*see also* killing and letting die)
Donaldson, S. 566n.7, 571n.23, 580n.47
Dorsey, D. 16, 93, 94n.2, 139n.3, 280–282, 281n.15, 312
Dougherty, T. 49, 280
Dowe, P. 636
Dreier, J. 3, 3n.6, 10, 16, 25, 25n.1, 29n.8, 29n.9, 30n.12, 31n.16, 33n.18, 39n.29, 50n.6, 51n.8, 56–57, 61, 67n.2, 85, 86, 86n.21, 139n.1, 276n.7, 279–280
Driver, J. 19, 249n.31, 403n.7, 409n.13, 425, 431n.13, 463–465, 468, 471, 617, 625–626, 629–630
Drummond, M. 594, 603
dualism of practical reason 383n.5
dual-ranking act-consequentialism. *See* consequentialism, dual-ranking act-
Duflo, E. 617
duty: (*see also* obligation)
  aid and 56
  beyond the call of 269, 272, 281–282, 499
  imperfect 299n.16
  perfect 299n.16
  prima facie 292, 328, 376, 407
  pro tanto 113, 122, 125, 132
  special 47, 49–51, 54
economics 20, 199n.8, 242, 248–249, 385, 387, 401, 404n.9, 416, 504n.8, 537, 538–539, 544, 555–579, 593–595, 597–598, 600, 602–603, 605
Edlin, A. 639
(p. 662) effective altruism 17, 19, 146, 531–546, 548–563, 566–567, 577, 579, 610, 641
egalitarianism 226, 379, 383, 385–386, 388–391, 398, 408, 499n.1, 566, 580, 601, 622
egoism 5, 6, 46n.1, 203, 257, 277n.10, 302, 321n.6, 381–382, 383n.5, 387, 408, 414
environment 19–20, 206–207, 459, 536, 545, 573, 576, 578, 583, 592–611
equality 369, 374–375, 381, 385, 442, 449, 469, 480, 499n.1, 616, 619–620, 622–623, 623n.10, 626
Eriksson, A. 637, 640
error theory 251
Estlund, D. 7
ethical egoism. *See* egoism, ethical
evaluative focal point 428–431, 436, 438

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

evaluator-relative. *See* agent-relative
excuses 165, 168, 170, 250, 296, 303, 306, 313, 320, 335, 456
experience machine 214, 214n.64
externalism 211, 211n.47, 213, 466
factory farms 414, 539, 574, 587, 639
family 48, 200, 222–223, 226, 240n.10, 241, 254, 260, 274, 405, 410, 451, 463, 519–520, 532, 545, 574, 619–621, 625
Feldman, F. 97n.7, 115n.6,143, 147n.15, 147n.17, 172, 204, 204n.27, 204n.30, 212n.49, 214n.59, 272–273, 316, 409n.13, 435n.23, 443n.3, 476n.5, 515n.5, 516n.6, 520n.18, 565n.3
feminism 20, 257, 257n.9, 272, 616–631
Fenton-Glynn, L. 639
Ferry, M. 271n.2, 271n.3, 279, 283–285, 284n.17
Finlay, S. 253n.1
Finn, H. 326n.9, 465
Fischer, B. 571n
Fischer, R. 609–610
fittingness 61–62, 285, 402, 404, 407, 410–411, 430, 501–503, 503n.6, 506, 511n.13
Flanagan, J. 254n.5, 617n.5
Flescher, A. M. 280
Fletcher, G. 204n.29, 205n.33, 205n.34, 442
Fleurbaey, M. 594, 601–603
Flores, A. R. 576n.42
Foot, P. 31n.15, 67, 185n.13, 203, 203n.24, 361, 361n.6, 368–370, 376, 376n.13, 424, 561, 561n.50
Forcehimes, A. 63n.24, 182n.9
Foucault, M. 585n.54
Frankena, W. 200, 200n.11, 453
Frankfurt, H. 37n.26, 164–165, 176
Franklin, A. 93, 93n.1, 95, 101, 107
Fraser, C. 10n.17
Fraser, D. 607
friendship 18, 186, 199, 225, 239n.5, 240n.10, 241–243, 245–248, 250, 261, 299n.17, 301, 367, 381, 402, 402n.5, 404–405, 404n.8, 408–409, 411–414, 414n.19, 449, 465, 552, 625
Frijda, N. 212–213, 213n.56
future generations 207, 253, 414, 577, 601–602
Gabriel, I. 541
Garcia, J. 52n.12
Garnett, T. 610
Garrett, D. 470n.6
Garvey, J. 643
Gauthier, D. 360
George the chemist 173–174, 176, 634–635, 649
Gesang, B. 643
Gibbard, A. 305n.29, 516n.7, 517n.12, 519nn.14–15
Gilbert, M. 521n.21, 524n.28
GiveWell 555, 558, 558n.38, 594
Glasgow, J. 20n.27, 116
global poverty. *See* poverty

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

global warming 312–313
Glover, J. 634, 639
Goble, L. 87n.32
Goldman, A. 114n.5, 118–120, 118nn.15–16, 118n.18, 119n.21, 120nn.22–23, 122n.27, 137n.43
Goldman, H. S. 115nn.6–7, 144n.12, 146n.13, 147n.15, 147n.17, 149–150 150n.22, 152, 305n.29, 337n.8, 514n.3, 515n.5
Goldstein, L. F. 618n.7
Goodin, R. E. 622, 630
goodness. *See* value
Graham, P. 318
Greaves, H. 19, 312, 409n.13, 409n.14, 423, 437n.24, 577n.24
(p. 663) Greene, J. 544–546
Greenspan, P. 147n.18
Griffin, J. 208n.39
Groff, Z. 580n.48
Grotius, H. 469
Gruen, L. 569, 569n.19, 570n.21, 571n.23, 571n.25, 573, 573n.29, 576n.38, 584n.52, 605
Gruzalski, B. 519n.16
Gunnemyr, M. 643
Gustafsson, J. 113n.1, 114n.2, 115nn.6–7, 147n.17, 150n.22, 156n.32, 565n.4
Haines, W. 98n.9
Hammerton, M. 46, 48n.3, 49, 51n.9, 57, 59, 59n.20, 63nn.25–26, 240n.9
Hanson, R. 570n.20
Hansson, S. O. 84, 87n.22
Hare, C. 2n.5, 15n.24, 116n.13, 344, 346n.3, 349n.4, 479n.11, 494n.53, 513n.1
Hare, R. M. 199, 199n.7, 411, 425, 437, 447n.4, 565, 565n.3, 568nn.15–16, 570n.20, 574
Harris, J. 231n.13, 253n.4
Harrison, J. 519n.16
Harrod, R. F. 457n.12, 519n.16
Hart, H. L. A. 96, 96n.4
Harwood, S. 271n.3
Hausman, D. 596
Hawkins, J. 205n.33, 206n.35
Heathwood, C. 200n.10, 211nn.46–47, 212n.49
Hedahl, M. 267n.24
hedonism 197, 199–200, 199n.5, 201n.13, 208, 210, 212n.49, 213–215, 214n.59, 241, 302, 370–371, 380, 557n.37, 565, 603, 628
Held, V. 623
Henne, P. 101n.14
Herculano-Houzel, S. 607, 609
Herzog, L. 536
Heyd, D. 271n.3, 272
Hiller, A. 603, 605, 643
Hills, A. 464, 470
Hilpinen, R. 85n.20
Hitchcock, C. R. 639
Hitler, A. 96
Hodgson, D. H. 297n.14, 424, 519n.13

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Holtug, N. 479n.10, 496n.57
Hooker, B. 4n.8, 9, 19, 84, 168n.8, 186, 186n.14, 188–190, 189n.16, 225n.5, 239n.5, 283, 303–306, 305n.29, 312, 426, 427n.6, 432, 441–442, 450, 555n.28, 561n.50
Honoré, A. M. 96, 96n.4
Horgan, T. 271n.2, 279n.12, 312n.1
Horty, J. 85
Horwich, P. 335
Howard-Snyder, F. 4n.7, 303n.24, 501, 501n.3, 509
Hsiung, W. 599, 610
Hubin, D. 544
Huemer, M. 492n.46
Hume, D. 212n.49, 290, 290n.3, 376, 376n.13
Hurka, T. 47n.1, 53n.13, 407, 431n.13, 469
Hurley, P. E. 3n.6, 25, 41, 86n.21, 240n.9, 302n.23, 443n.2, 543
Hursthouse, R. 226n.8, 431n.13, 470
hybridism 140–141, 151, 157–159
hybrid theory 298n.15, 403, 406–408, 416
Hyde, D. 394n.10
ideal code 4, 4n.8, 9, 10n.18, 186, 189n.16, 225n.5
idealization 53, 186, 188, 234n.17, 380, 407, 514–515, 517n.8, 518, 519n.14, 523, 525n.31, 575, 583, 622
impartiality 9, 36, 62, 64, 197, 203–204, 239, 243–244, 247, 248, 255, 258–261, 276, 310n.21, 322–324, 326, 359–360, 378, 381, 405–406, 410, 416, 437, 442–443, 452–453, 454, 457, 457n.11, 463–464, 502, 533, 541–542, 552, 566, 580, 625–626, 629
imperfect duties. *See* duty, imperfect
inclusivism 333–334
incommensurability 201, 201nn.15–17, 359, 367–376
incomparability 201, 201nn.15–17, 202, 368–376, 369n.12
indeterminacy 15, 15nn.24–25, 135n.40, 344–357, 374, 391, 394, 394n.10, 397, 513
indirect consequentialism. *See* consequentialism, direct vs. indirect
(p. 664) inequality 226, 495, 538, 550, 556n.31, 602–603, 611, 618, 623
inheritance. *See* normative, inheritance
internalism 211n.47, 213
intuitionism 269, 292n.5, 557
intuitions 59, 59n.20, 61, 64, 129, 132–133, 167–168, 176, 199–200, 202, 204, 211, 223, 225n.2, 227–229, 231, 232–233, 240, 250–251, 270, 273, 278, 282, 286, 293–297, 294n.9, 300, 303, 306, 308, 333, 391, 431, 457, 457n.12, 492n.47, 500, 541–542, 544–546, 551, 557n.37, 562, 581, 621, 625, 629–630, 635
Jackson, F. 16, 20n.27, 84n.18, 115n.7, 140n.7, 148n.19, 151n.27, 171, 171n.14, 240n.8, 313–321, 313n.2, 314n.3, 331, 332n.1, 334n.5, 337–338n.8, 405, 408, 434, 444, 466, 514n.4, 635, 639, 644, 646
Jaeggi, R. 401n.2
Jaggar, A. 631
Jamieson, D. 585, 605
Jarvis, L. 599, 610
Jefferson, A. 166
Jeffrey, R. C. 434n.21
Jeske, D. 17, 238, 240n.10, 241n.11, 242n.12, 249n.32, 448, 460nn.45–46, 540, 559–560

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Jim in the jungle 173–176
Johnson, B. 344
Johnson, C. 55n.14, 59n.19
John, T. M. 19–20, 464, 543, 546
Jollimore, T. 241n.11, 250, 250n.34
Kagan, S. 180, 185, 190–191, 190n.18, 272–273, 342, 381, 382n.3, 409n.13, 428, 507, 523n.25, 554, 554n.25, 570n.21, 593, 605, 630, 639–640, 646–648
Kahn, B. 644
Kahneman, D. 601
Kain, P. 401n.2
Kamm, F. M. 55n.14, 56, 98–99, 99n.11, 102, 174n.16
Kant, I. 2–3, 6, 12n.21, 13, 32, 116, 170, 321n.6, 328, 376
Kantianism 12–16, 12n.21, 27, 27n.5, 31–33, 75n.11, 116n.9, 186, 269, 328, 362, 508, 561, 566, 626, 630
Kantsequentialism 2–3, 5–6
Kapur, N. B. 245, 245n.22, 412
Kavka, G. 490n.40, 491n.42
Kawall, J. 271n.3
Khader, S. J. 627, 630
Kiesewetter, B. 121n.26, 147n.15
killing and letting die 507–508, 507n.10
Kingston, E. 643
Klein, C. 607
Kment, B. 639
Kolstad, C. 597–598, 600, 603
Korsgaard, C. M. 566n.7
Kotarbinski, T. 269n.1
Kraut, R. 393n.8, 398n.13
Kringelbach, M. 212nn.54–55, 213n.56
Krishna, N. 540
Kuhse, H. 625–626
Kumar, V. 19, 531, 543–545, 563n.53
Kunst, J. R. 574, 574n.34
Kymlicka, W. 566n.7, 571n.23, 580n.47, 623n.10
labor injustice 20, 639–642, 648, 650
Labukt, I. 211n.45
Lam, D. 398n.13, 595
Lang, G. 273, 287, 501n.4
Lawford-Smith, H. 20, 168n.8, 634,639–643, 647
Lawlor, R. 501–502
Lazari-Radek, K. de 17, 169, 197, 203n.22, 209n.41, 226n.10, 273, 447n.4, 477n.7, 501n.2
Lazar, S. 13
Lenman, J. 96n.5, 558, 558n.41
Leopold, D. 401–402, 401n.2
Lewis, D. 15n.24, 88n.23, 140n.7, 149n.21, 346nn.2–3, 350–352, 351n.6, 353n.8, 513n.1, 636–638
Lewis, G. 556
Lichtenberg, J. 19, 533, 546, 548, 549n.6, 561n.49
Li, H. L. 391n.7, 398n.13

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

list theory. *See* objective list theory
Lockhardt, T. 326
logic of the larder 569–570, 587
logic of the logger 579, 580, 587
Lombard, L. B. 116n.11
Lopez, T. 55n.14
Lord, E. 181n.5, 468n.4
(p. 665) Loughnan, S. 573n.32, 574n.33
Louise, J. 25n.1, 29n.9, 30n.14, 51n.8, 55n.14, 59n.19, 179, 410n.16
Lyons, D. 191n.20, 305n.29, 459, 561n.50
MacAskill, W. 17, 531, 533–537, 539–541, 546, 548–549, 551–552, 552n.15, 554, 556, 556nn.33–34, 559, 559n.42, 566n.8, 567n.10, 577n.44, 593n.1, 641
MacFarquhar, L. 552n.16, 557n.35, 559, 559n.44
Mackenzie, C. 624
Mackie, J. L. 195n.23, 342
Maguire, B. 18, 20n.27, 174n.17, 401–402, 402n.4, 404, 406n.10, 407–409
Marcus, R. B. 267n.24, 402, 432n.17
Markovits, J. 404, 470
marriage 117–122, 298–299, 618–622, 641
Mason, E. 18, 162, 168n.9, 170, 173, 175, 247, 247nn.26–27, 411, 413, 449n.5, 468
Matheny, G. 571, 571n.24
maxim 116, 328, 359–360
maximal alternatives. *See* actions, maximal and nonmaximal
maximalism 84, 136n.42, 140n.4, 143–144, 150n.22, 156
maximally specific act-sets. *See* actions, maximal and nonmaximal
maximizing 499–501, 499n.1, 505, 506, 508–511, 524, 533, 540, 550, 554–555, 555n.28, 561, 563, 565–566, 565n.4, 568, 571, 578, 593, 602, 605, 616n.1, 623, 629–630
McElwee, B. 179, 181n.6, 234n.15, 273, 285, 287, 500, 502, 502n.5
McGeer, V. 165n.5, 166, 166n.6
McGrath, S. 99n.12, 101n.14, 641
McKerlie, D. 389n.6
McLeod, C. 406n.10
McLeod, O. 624
McMahan, J. 481n.19, 537–538, 540, 550, 551n.12, 607n.5
McNamara, P. 85n.20, 271n.3
McNaughton, D. 48n.3, 303n.24, 568n.16
McShane, K. 605
McTaggart, J. M. E. 392, 392n.8
meat eating 261, 376, 570n.21, 572, 574–575, 578, 581, 593, 617
meat paradox 573
Mendola, J. 7n.13, 19, 20n.27, 513, 513n.1, 517n.10, 521nn.21–24, 524n.27
Menzies, P. 315, 317–318, 639
mere addition 482–483, 484n.23, 486, 488, 492n.45, 493
mere addition paradox 474–475, 482, 484–485, 487, 491–495
mere addition principles 475, 482–483, 485, 570, 571n.23
Meyers, D. T. 257n.9
Mikati, I. 602
Mikhail, J. 118n.15

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Mill, H. T. 619n.8
Mill, J. 17
Mill, J. S. 17, 27, 199, 214, 214n.60, 226n.11, 242, 242n.13, 283, 290, 290n.2, 340–342, 392, 392n.9, 396, 425, 447n.4, 464, 531, 553, 553n.17, 553n.19, 553n.21, 554, 564, 618–622, 618n.7, 627
Milne, P. 67n.2
minimal decency 504, 506, 510–511
Mitchell-Brody, M. 564n.1, 585n.54
Moore, G. E. 47, 97, 97n.6, 98n.8, 198n.4, 199, 199n.9, 203, 206, 208, 208n.38, 312n.1, 313, 316, 333n.2, 344n.1, 554, 554n.23
Moore, M. 382n.3
moral:
- actualism 479–482, 479n.11, 481n.17, 494, 494n.53 (*see also* actualism)
- circle expansion 577
- dilemmas. *See* delimmas
- possibilism 479–481, 488, 491, 493–494
- properties 121–123, 121n.24, 132, 134–135, 135n.39, 136n.42
- rationalism 231, 234, 234n.16, 265
- reasoning 617, 619, 621–625, 629
- reasons 4, 28, 36–38, 94, 96, 111, 166n.6, 182, 194, 256, 265–267, 277–282, 284, 285–286, 295n.11, 296, 358, 361n.6, 364, 381, 501–502, 506, 509–510, 568, 570, 580, 648
- responsibility. *See* responsibility, moral
- sentimentalism 470n.6, 506–508, 510
- status 19–20, 32–33, 113, 119n.19, 123, 135, 240, 255, 271, 277, 280, 285–286, 344–345, 349–350, 354–357, 478–479, 492, 494, 564, 572, 574–575, 581, 617, 644
- uncertainty. *See* uncertainty, moral

(p. 666) Moran, R. 42n.31
Morgan-Knapp, C. 643
Morris, R. 163n.1
motivating reasons. *See* reasons, motivating
Mozi 10n.7
Mulgan, T. 261n.13, 275, 452n.9, 457n.12, 496n.57, 505, 507
Murphy, L. 226n.9, 235n.18, 261n.13, 500
Nagel, T. 36, 36n.23, 39, 41, 202n.18, 360, 381, 382n.3, 383, 385, 388–389, 625n.12
Nair, G. S. 10n.19, 50n.5, 67, 68n.3, 72
Narveson, J. 290n.2, 291, 292, 297n.14, 479, 479n.12, 481
Nebel, J. 391n.7
Nefsky, J. 593, 639, 641, 647
NESS conditions 645, 648–649
Newcomb’s problem 517n.12
Ng, Y.-K. 486n.32, 492, 492n.47, 493, 580, 605
Nilsson, M. 604
Noddings, N. 624
Nolt, J. 593, 643
nonconsequentialism 3, 10, 12, 16–18, 25–28, 30–33, 38, 62, 67–69, 71, 85, 116n.9, 164, 166, 169, 172–173, 194, 197, 227, 229, 235, 244, 293–294, 296, 302, 327–328, 360–361, 364, 407, 411, 414, 416, 438, 441, 457–458, 500, 503, 543, 565, 568–569, 571, 572, 578, 586–587, 630
non-human animals. *See* animals
nonidentity problem 475n.4, 490–491, 492n.47

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

nonmoral reasons 265–267, 277–279, 502, 506, 510
Norcross, A. 4n.9, 16, 139n.3, 184, 225n.6, 247n.27, 272–273, 287, 293–294, 293n.8, 336n.10, 358, 358n.4, 361n.7, 362n.8, 364n.9, 411, 500, 501–504, 502n.5, 509
Nordhaus, W. 598, 602
Norlock, K. 631
normative:
authority 407–408
inheritance 74–78, 82–85, 87, 89–90, 121–122, 121n.26, 130–131, 134, 143
invariance 494
properties 53, 125, 128–133, 394n.10, 508, 544
uncertainty. *See* uncerntaintly, normative
Norwood, F. B. 599–610
Nozick, R. 214, 214n.61, 214n.64, 327, 360, 382n.3, 383–384, 388, 517n.12, 523
Nussbaum, M. 617, 627–628
Nye, A. 639
Nye, H. 38n.28
Oakley, J. 173n.15, 246, 246n.24, 247, 247n.25, 412–413
objective list theory 208–210, 213, 225, 442, 603–604
objective ought. *See* ought, objective
objective rationality. *See* rationality, objective
objective reasons. *See* reasons, objective
objectivism 116n.13, 171n.13, 314, 320, 333–334, 333n.2, 334n.5, 467
obligations:
conditional 147–148
conflicting 153, 297
dependent 143–144, 146, 151
nondependent 143–144, 146, 150, 155, 159
possibilist moral 157–158
promissory 18, 289–292, 296–298, 300, 302–307
pro tanto 296, 297, 300, 304, 306, 307
special 10, 260–261, 267, 541–543, 546
unconditional 147–148
Oddie, G. 67n.2, 203n.25, 315–318
Okin, S. M. 619
Olkowicz, S. 607
omissions 99–101, 103, 295, 566, 636
omniscience 316, 316n.12
omnism 133–134, 140n.4, 144, 150
O’Neill, J. 605
options:
agent-centered 10
c-options 339–341
deontic 68–70, 79
maximal. *See* act, maximal and nonmaximal
ordinary morality. *See* commonsense morality
Ord, T. 17, 531, 548, 566
Ostrom, E. 595, 600
Otte, M. J.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

ought:
actualist practical 158
(p. 667) 'can' and whether it's implied by 155, 428n.8, 433n.19
just plain 406, 406n.10
most reason and 19, 498, 510–511
nonobjective 317
objective 33, 35n.20, 317, 320
prospective 334n.6
rational 34–35
outcome of an act (*see also* prospect of an act)
overdetermination 636–638, 645, 647–648
overridingness 244, 247, 305, 407, 450, 457–458
Oxfam 278, 341
Palmer, C. 605
paradox of supererogation. *See* supererogation, paradox of
pareto 72, 485–487, 489–494, 597
Parfit, D. 37n.25, 48, 55, 55n.16, 58–59, 59n.18, 63–64, 63n.25, 168–169, 180n.3, 184, 188, 190, 199, 200n.10, 208–209, 225n.3, 313, 391, 391n.7, 396, 396n.11, 397n.12, 409n.13, 424–425, 424n.1, 428, 431, 432n.15, 442, 442n.3, 445, 459, 474n.1, 475n.4, 482–485, 488, 490–492, 494n.53, 503, 508, 517n.9, 565n.3, 639, 644–646, 649
Pargetter, R. 84n.18, 115n.7, 148n.19, 151n.27, 337n.8, 434, 514n.4
Parsons, J. 479n.13
partiality 17, 64, 79, 120, 203–204, 233–234, 238–251, 260–261, 264, 277, 296, 299n.16, 300, 301n.21, 322, 326, 332, 381, 405, 407–408, 410, 416, 425, 429, 463, 465–466, 502, 533, 540–541, 545, 559–561, 593–594, 625–626, 641
patriarchy 257, 618, 620–621, 627
Pearl, J. 88n.23
Pellegrino, G. 643
people:
actual 478–480
merely possible 479, 492–493
perfect duty. *See* duty, perfect
perfectionism 285, 621, 628–629
permissions 249, 249n.33, 258, 260, 263–264, 267, 272, 282n.3, 386, 407, 452, 630
Peterson, M. 25n.1, 29n.9, 30n.14
Pettit, P. 47n.2, 63, 64, 1972, 197n.1, 198n.2, 245n.21, 275–276, 300–301, 301n.21, 303n.24, 382, 406, 408n.13, 411, 425, 449n.5, 505
philosophical radicalism 17
Piazza, J. 576n.38
Pinillos, Á. 101n.14
Pinkert, F. 523n.26, 525n.30, 526n.32, 641
Plato 392, 531
pleasures:
bodily 212
emotional 212, 243
higher 392, 392n.9
intellectual 211–212, 214
lower 214, 392, 396

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

resonates requirement and 210–213
Podgorski, A. 188
Pollock, J. L. 149n.21
pollution 207, 367, 523–524, 594, 596–598, 600, 602, 611, 644
population axiology. *See* axiology
pornography 621
Portmore, D. W. 1–3, 4n.7, 7n.13, 25n.1, 26n.2, 29n.8, 30, 30n.12, 31n.15, 33n.17, 35n.20, 35n.21, 36, 36n.23, 39n.29, 51n.8, 52n.12, 55n.14, 57, 59n.20, 61n.23, 67nn.1–2, 76n.13, 84, 113n.1, 114nn.2–4, 115nn.6–8, 116n.9, 119n.20, 121n.26, 123n.28, 133, 135n.42, 137n.43, 139n.1, 143, 143n.11, 144, 144n.12, 146n.13, 150, 150n.22, 151n.27, 152, 156, 159n.33, 179, 182n.9, 184, 192, 194, 195n.23, 234n.16, 237n.19, 251n.38, 265, 265nn.19–22, 266, 271nn.2–3, 277–279, 281n.15, 303n.24, 308n.30, 312, 312n.1, 317, 333n.4, 339n.9, 382n.4, 398n.13, 401n.1, 407, 434n.21, 439, 441–442, 442n.3, 443n.2, 459, 460n.15, 502, 509n.11, 511n.14, 515n.5, 527n.33, 546, 555n.28, 557n.37, 563n.53, 564n.1, 566n.5, 593, 593n.1, 624n.11, 631
position-relative. See agent-relative
positive act. *See* acts, positive
positive duty. *See* duty, positive
Posner, R. A. 570n.20
possibilism 74n.10, 81–84, 86–87, 89–91, 139–159, 188n.15, 434–435, 572n.26
(p. 668) Postow, B. C. 520n.19
poverty 20, 69, 275, 414, 500, 505, 531, 537, 539, 542, 546, 548–550, 552–553, 555–557, 560–561, 639, 641–642, 646, 648–650
Powers, M. 239n.5, 449n.5
practical guidance. *See* action, guidance
practical reason, reasons, and reasoning 14, 14n.23, 31–32, 35, 93, 158, 201, 264n.17, 275, 276, 383n.5, 405, 506
praiseworthiness 121n.24, 170–172, 255, 555
Prawitz, D. 114nn.2–4, 115n.6, 115n.7, 118n.15, 142, 142n.9, 144
preference satisfaction theory. *See* desire satisfactionism
principle of moral harmony 516, 518
principle of normative invariance. *See* normative, invariance
prioritarianism 389–390, 442, 601
prisoner’s dilemma 63, 73n.8, 516
probabilities:
conditional 350–351, 355–357
epistemic 204, 315–316, 322, 336
objective 11, 11n.20, 15, 15n.24
subjective 314, 336, 342
Professor Procrastinate 434–435
promises 28–29, 58, 119, 130, 183, 191–193, 233, 240, 289–308, 327–328, 332, 447–448, 450, 458, 518
prospectivism 171, 171n.13, 344n.6
prospects of an act 2n.5, 11–15, 75, 318–319, 325–326 (*see also* outcome of an act)
prudence 93–95, 116–117, 140, 225, 225n.7, 226n.8, 254, 281, 299, 338, 359–360, 368, 383, 387, 397
prudential aggregation. *See* aggregation, prudential
prudential reasons. *See* reasons, prudential
prudential value. *See* value, prudential

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

public policy 20, 415, 592–611
Pummer, T. 391n.7, 397n.12, 398n.13
punishment 167, 412, 454, 513, 576
QALYs 534, 538
Quinn, W. S. 362n.8, 639
Rabinowicz, W. 198n.4, 269n.1, 494, 494n.52
Rachels, J. 253n.3
Rachels, S. 362n.8, 391n.7, 397n.12, 477n.9
racism 402, 414, 531, 555, 627
Railton, P. 27n.4, 203n.26, 204–206, 205n.31, 215, 225n.4, 243–247, 244nn.16–17, 244n.18, 246n.23, 301, 411, 414, 424–425, 428, 431, 465–466, 471, 500, 501n.3, 561n.51
Ramsey, F. P. 336
rankings:
- agent-relative 28, 35, 46, 50–51, 54–55, 416
- location-relative 56–57, 59
- patient-relative 57–58
- position-relative 46–47
- time-relative 46, 55
- world-relative 28n.6, 56–57, 59

Raphael, D. D. 95–96, 96n.3
rational egoism. *See* egoism
rationalism. *See* moral rationalism
rationalism rejoinder 234–235
rationality:
- decision-theoretical 317
- instrumental 205–206
- practical 250, 276, 278, 499

Rawls, J. 303, 303n.25, 360, 380, 382–389, 411–412, 452n.8, 457n.12, 601
Raz, J. 264, 264n.17, 369, 369n.12, 375
reactive attitudes. *See* attitudes, reactive
reasons:
- action and 4, 18, 32, 36n.22, 37, 179–195, 231, 250, 256, 266, 273, 502, 503n.6, 506, 509
- agent-neutral. *See* agent-neutral, reasons
- agent-relative. *See* agent-relative, reasons
- enticing 279
- epistemic 315
- moral. *See* moral reasons
- morally relevant 32, 277
- motivating 180, 180n.4, 404–405
- nonmoral. *See* nonmoral reasons
- normative 180, 180n.2, 180n.3, 182, 503n.6, 507
- objective 111
- pattern-based 187–190, 188n.15, 189n.17, 192
- practical. *See* practical reason, reasons, and reasoning
- prudential 182, 387, 502
- subjective 10n.17, 22

receptacles 8, 464
redundancy worry 230, 235–236

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

(p. 669) Reese, J. 569n.18
reflective equilibrium 168, 457, 457n.12, 459–460
Regan, D. 7n.11, 7n.13, 52n.11, 60n.22, 63n.24, 64n.27, 313n.2, 332n.1, 334n.6, 516n.7, 517n.9, 524n.29, 525n.31
relationships. *See* special relationships
relative value theory. *See* consequentialism, agent-relative
Rendall, M. 643
repugnant conclusion 379, 391–392, 396–398, 396n.11, 397n.12, 475n.4, 483, 491–492, 492nn. 45–46, 601, 608–610
requirements (*see also* duty):
  etiquettical 281
  legal 281
  moral 228, 231–232, 240, 279–282, 285, 293n.8, 433, 436, 562, 629
  prudential 281
  rational 281–282
resonance requirement 205–206, 208, 210–211, 213, 215
responsibility:
  consequentialist accounts of 162–168
  individual 174, 515
  moral 18, 162–177, 638–639
  proportionality and 638
responsibility constraint on rightness 171, 174, 176–177
restrictions. *See* constraints
Ridge, M. 186
rights 20, 27, 27n.5, 46, 55, 63, 71–72, 77, 117, 190–191, 194, 227, 232, 296–297, 304n.28, 360, 381, 383–384, 493, 533, 543–544, 557, 560, 562, 564, 568–569, 571, 575–576, 579, 611, 616n.1, 617–619, 623–624, 628–630, 640
Roberts, M. A. 19, 184, 474, 476n.5, 479n.10, 479n.11, 480n.16, 481nn.18–19, 487n.33, 489n.38, 490n.40, 491n.43
Ross, J. 84, 147n.17, 151n.29, 153–154, 159, 191, 326, 515n.5
Ross, W. D. 202, 250, 291–292, 294, 300n.13, 310, 407
Rothgerber, H. 575n.36
Rowlands, M. 566n.7
rule consequentialism. *See* consequentialism, rule
rules:
  compliance with 185–190, 429, 458–459
  embedding 185–189, 190n.18
  internalization of 165, 283, 304–306, 432, 449, 451–452, 454
  part of a criterion of rightness and 452–455
  part of a decision procedure and 445–452
  publicity and 427n.7, 452, 452n.8, 561n.51
  of thumb 243, 245, 294, 341, 424
rule utilitarianism. *See* utilitarianism, rule
Russell, P. 470n.6
Ryle, G. 316n.4
Sachs, B. 31n.16
Salt, H. S. 570n.20
Sandberg, J. 643

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Sandler, R. 600, 602n.3, 603, 605
Sarkar, S. 605
Sartorio, C. 636, 639, 641
satisficing consequentialism. *See* consequentialism, satisficing
Sauer, H. 31n.15
Scanlon, T. M. 35n.21, 163n.2, 180n.3, 202n.20, 295n.11, 298n.15, 302, 361–362, 361n.7, 385–386, 389, 390, 405, 447n.9
Schacht, R. 401, 401n.2
Schaffer, J. 639
Scheffler, S. 10, 27n.5, 47n.2, 68n.4, 173–175, 183n.9, 190, 190n.19, 249n.33, 261n.13, 262n.15, 263, 263n.16, 276, 406, 442, 499
Schlick, M. 162–163
Schlottmann, C. 567n.12, 570n.21, 575n.37
Schmidt, A. 163n.1, 169n.10
Schmidt, D. 605
Schroeder, A. 29n.10, 30n.11, 37n.27
Schroeder, M. 29n.8, 52n.11, 60–62, 60n.22, 67n.2, 179, 198n.4, 317, 404–405, 407n.12
Schroeder, T. 468n.5
Schubert, S. 586
Schwenkenbecher, A. 643
Scovronick, N. 601
Sebo, J. 19–20, 538, 564, 567n.12, 569n.17, 570n.21, 575n.37, 585n.54, 609
securitism 140–141, 149–157, 159, 434n.21
self-defeat 19, 63, 374, 423, 424n.1, 425, 436, 438–439, 514–519, 523, 526, 621
Seligman, M. 412
(p. 670) self-other asymmetry. *See* asymmetry, self-other
self-sacrifice 259n.10, 259n.11, 448, 451, 457, 461
Sen, A. 10n.17, 25n.1, 37n.27, 46–47, 51n.8, 60n.21, 64n.27, 199n.5, 397, 617
separateness of persons. *See* consequentialism, act, separateness of persons objection and
Sepielli, A. 320n.5, 326, 327n.11
Setiya, K. 49
sexism 531, 622, 627
Shahar, D. 605
Shaver, R. 211n.46
Sherwin, S. 623–625
Shizgal, P. 212, 212n.55
Shrader-Frechette, K. 602
Shriver, A. 607
Sider, T. 259, 259nn.10–11, 277n.10
Sidgwick, H. 47n.1, 162, 169, 198n.4, 199, 199n.6, 203, 203n.21, 203n.23, 211n.45, 212n.49, 213, 213n.57, 215, 226n.10, 242–243, 273, 290n.2, 293, 340–342, 359–360, 380n.2, 383, 383n.5, 384, 409, 425, 447n.4, 449n.5, 533, 553–554, 553n.18, 554n.22, 561n.51, 563n.52, 564–565, 565n.3, 568n.16
Silverstein, M. 214, 214n.65
Simon, H. A. 493, 504n.8
Singer, P. 17, 169, 209n.41, 226n.10, 242, 253n.2, 255n.6, 273, 290n.2, 292, 295, 297n.14, 447n.4, 477n.7, 481n.19, 499, 501n.2, 505, 511n.14, 531–533, 537–538, 540–546, 548–554, 548n.3,

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

549n.4, 556–557, 559n.43, 563n.53, 564, 566, 569n.17, 570nn.20–21, 571n.23, 572, 580n.47, 593, 605, 609, 625, 641
Sinhababu, N. 504, 565n.4
Sinnott-Armstrong, W. 5, 283, 291n.4, 432n.17, 554n.27, 566n.5, 639, 643
Skorupski, J. 467n.3
slavery 221–222, 226, 326n.9, 465, 562, 620–621
Slote, M. 84, 239n.6, 254n.5, 274–275, 452n.9, 470n.6, 471, 499n.1, 500–505
Smart, J. J. C. 162–163, 170, 333n.2, 341, 424–425, 427, 634, 642
Smith, H. M. 16, 20n.27, 113, 126n.31, 172, 184, 186, 312, 337n.8, 340 (*see also* Goldman, H. S.)
Smith, M. 25n.1, 29n.8, 35, 35n.20, 37n.25, 37n.26, 51n.8, 52, 52n.12, 53, 53n.13, 61–62, 302, 312n.1, 318, 325n.8, 409n.13
Smuts, A. 211n.45
Snedegar, J. 84n.18, 181n.8, 183n.11
Sobel, D. 17, 194, 221, 272, 299n.18, 342, 448, 500, 544
Sobel, J. H. 115n.7, 144, 147n.17, 150, 150n.23, 337
Socrates 392
sorites 379, 392–396, 398
special duties. *See* obligation, special
special obligations. *See* obligation, special
special relationships 49, 51, 238–241, 240n.10, 256, 260–261, 560
speciesism 322–325, 575, 582–583, 587
Spiekermann, K. 643
Srinivasan, A. 536, 551n.13, 556n.34
Stalnaker, R. 15n.24, 140n.7, 149n.21, 346n.2, 346n.3
Star, D. 180n.2
Stephen, L. 570n.20
Stiglitz, J. 602
Stocker, M. 175n.18, 200n.12, 202, 202n.19, 424, 464
strategic response 242–245
Strawson, P. F. 164–165
subjective ought. *See* ought, subjective
sufficientarianism 442
Suikkanen, J. 62, 279, 286
Sumner, L. W. 39n.29, 208n.39, 211n.45, 211n.47
Sunstein, C. 594, 599–600, 610
supererogation 17, 121, 131, 133, 158, 197, 269–287, 499, 505–506
Superson, A. 627, 630
Swanton, C. 470n.6
Talbot, B. 641
Tankard, M. E. 576n.42
Tännsjö, T. 411, 492n.46, 609
Taurek, J. 361, 361n.6
Taylor, J. 470n.6
Taylor, S. 531, 585n.54
Tedesco, M. 411, 413
(p. 671) teleological conception of practical reasons 14, 14n.23, 35n.21
teleology 60
telic equivalence thesis 32

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Temkin, L. 57, 362n.8, 397n.12, 475n.3, 477n.9, 485nn.55–56, 495, 495n.54, 623
Tenenbaum, S. 31n.15
Terlazzo, R. 630
Tessman, L. 624
theological consequentialist voluntarism 1n.3
Thompson, M. 42n.31
Thomson, J. J. 116n.11, 361, 361n.6, 622n.9
thresholds 15, 51n.7, 55n.15, 84, 398, 442, 498, 505, 609, 640, 642, 646–648
Tilly, C. 539
Timmerman, T. 17, 74n.10, 114n.4, 115, 133, 133n.37, 139, 140n.6, 143, 146n.14, 147n.17, 148n.19, 149, 150n.25, 151, 151nn.27–29, 158, 340, 444, 564n.1, 572n.26
Timmons, M. 271n.2, 279n.12, 312n.1
Todd, B. 548
Tomasik, B. 567n.13, 579–580, 582–584
too demanding objection. *See* demandingness objection
total principle 477–481, 488, 491, 494
traditional act-consequentialism. *See* consequentialism, traditional act
transitivity 57, 71, 106n.19, 120, 326n.10, 361, 362, 364, 369, 375–376, 379, 382, 395–398, 484, 494–495
trolley problem 67, 503n.7, 622
Tubman, H. 221–224, 226–227
Tuckwell, W. 20, 168n.8
Tuomela, R. 521n.20
Turri, J. 505
Tye, M. 607
Ullmann-Margalit, E. 181n.8
uncertainty:
  empirical 171n.13, 294, 312–322, 325, 327, 327n.11, 444, 518, 558, 567, 586–587, 595, 640
  epistemic 513
  evaluative 320n.5, 322–327
  factual. See uncertainty, empirical
  moral 294, 326–328
  normative 171, 171n.13, 593n.1, 609–611
  rational 320n.5, 326–327
Unger, P. 550–551, 551n.11
UNICEF 117 – 118, 120, 122, 123
Urmson, J. O. 269–271, 281, 285, 304n.26, 457n.12
utilitarianism:
  act 145, 191, 295–296, 358, 425, 447n.4, 451, 501n.3, 650
  against 142n.8, 215, 250, 292n.5, 295, 297n.14
  agent-neutrality and 382, 397
  classical 5–10, 46–47, 199, 203, 379, 380n.2, 386, 463, 514, 531, 533, 565–566, 568, 570, 580
  decision procedure and 243
  demandingness of 299, 549
  effective altruism and 532–535, 540–542, 553–554
  egoistically adjusted 277n.10
  ethics of fantasy and 342
  feminism and 618–619, 621, 623, 625–628

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

friendship and 243
global 425
hedonism and 214, 215, 380n.1, 380n.2
motive 257n.8, 471
promises and 191, 250, 290n.3, 291, 292n.5, 292n.6, 294, 297–299, 299n.16, 303–304, 307
rule 165, 303, 361n.6, 496n.57
scientific 542–545
self-evidence and 292, 292n.5
self-refute and 297n.14
separateness of persons and 360, 378–379, 382–390
supererogation and 269n.1
virtue and 464
welfare 32
value:
actual 323, 374, 445
aesthetic 596, 598
agent-neutral. *See* agent-neutral, values
agent-relative. *See* agent-relative, values
anthropocentric 599n.2
deontic. See deontic, values
impartial spectator and 62
(p. 672) impersonal 192, 381
instrumental 207, 214, 242, 367, 442, 543–544
intrinsic 36n.22, 140n.5, 145, 198–210, 215, 239–243, 245, 249, 296, 320, 322–324, 370, 443–444, 465n.2, 467, 563, 624
lexical priority and 77, 139n.3, 396
monism 27, 198–199, 199n.5, 200, 202
Moorean account of 53
moral 36, 38, 46n.1, 52, 60, 171, 214, 370, 376, 443, 444, 559, 573
non-moral 301n.20
pluralism 199–200, 202
promoting versus honoring 197, 301, 301n.21, 382
prudential 46n.1, 52
ultimate 203, 605
Vance, C. 643
Vanderheiden, A. 643
van Inwagen, P. 104, 104n.18
van Norden, B. 470n.6
Vargas, M. 166, 166n.6, 166n.7, 169
Velleman, J. D. 351n.5
Vessel, J.-P. 15n.24, 147n.15, 151, 271n.3, 271n.4, 277n.10
virtue:
Aristotelian accounts of 465
consequentialism and 463–471
ethics 31–33, 254n.5, 380, 431, 463–471
modal fragility and 466–467, 467n.3, 469
motivation and 328, 471
neo-Aristotelian accounts of 33

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

theory of 328, 431n.13, 464–470, 470n.6, 635
voting 20, 574, 582, 618, 639–641, 648, 650
Wallace, R. J. 163n.2, 298n.15
Walsh, M. B. 630
Wasserman, D. T. 184, 491n.43
Waters, C. K. 639
Watson, G. 380
Watson, L. 617n.5
Way, J. 181n.5, 403
Weathers, S. 538
Wedgwood, R. 84n.19, 148n.19, 151n.29
Weijers, D. 214n.64
Weinberg, R. 479n.13
Weintraub, R. 116n.11
welfarism 47, 202–208, 358, 566, 600–601, 606, 610–611
Weslake, B. 639
Westphal, F. 142, 142n.9
Whiting, D. 181
Wiblin, R. 567n.11
Wierzbicka, A. 212
Wiggins, D. 33n.19
Wiland, E. 568n.16
Wilcox, W. H. 246n.24
Williams, Bekka 522n.23
Williams, Bernard 67, 173–176, 201, 203, 250n.37, 277n.9, 383, 404, 414n.20, 432n.17, 433n.19, 436–438, 463, 463n.1, 540–542, 560n.47, 626n.13, 634, 642, 649
Williamson, J. 639
Wise, J. 552, 552n.16, 559n.42
Wolf, S. 277n.9, 451n.6, 459–460
Wong, K. 606, 607n.5, 610, 615
Wood, A. 401n.2, 402
Woodard, C. 18, 20n.27, 147n.17, 179, 181n.5, 186, 187, 188n.15, 250n.25, 442, 444, 447n.4, 459
Woodcock, S. 248, 248n.28
Woolf, S. 204n.29
Woollard, F. 233, 234n.15, 500
Wright, R. 645–646
Yuracko, K. A. 628–630
Zellner, H. 142–143
Zimmerman, M. J. 74 –78, 80, 82, 84n.18, 140n.5, 147nn.15–16, 147n.18, 151n.27, 151n.29, 171n.13, 312n.1, 314, 316, 316n.4, 320–326, 334n.5, 336n.7
Zuehl, J. 13n.22

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).