

OXFORD



THE COSMOS OF DUTY

Henry Sidgwick's *Methods of Ethics*

ROGER CRISP

THE COSMOS OF DUTY

THE COSMOS
OF
DUTY

HENRY SIDGWICK'S *METHODS*
OF ETHICS

ROGER CRISP

CLARENDON PRESS • OXFORD

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

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

© Roger Crisp 2015

The moral rights of the author have been asserted

First Edition published in 2015

Impression: 1

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 licence or under terms agreed with the appropriate reprographics
rights organization. Enquiries 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

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

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2014956615

ISBN 978-0-19-871635-8

Printed and Bound by
CPI Group (UK) Ltd, Croydon, CRO 4YY

Links to third party websites are provided by Oxford in good faith and
for information only. Oxford disclaims any responsibility for the materials
contained in any third party website referenced in this work.

For Harriet and Elizabeth

Preface

As Socrates says (Plato 2003: 352d6–7), how we should live our lives is no ordinary question. Indeed, it is prior to any other practical question we might ask, and the history of philosophical ethics—in so far as it is practical—can be seen as a series of attempts to answer it. Why should we study those attempts? They might be of purely historical interest, of course. But from the point of view of philosophical ethics itself, they matter primarily in so far as they provide resources for answering the Socratic question.

Henry Sidgwick's *Methods of Ethics* might seem an especially strong candidate for such study. C. D. Broad famously said that it seemed to him 'the best treatise on moral theory that has even been written' (1930: 143), an opinion recently echoed by Derek Parfit (2011: xxiii). Parfit claims—rightly, I believe—that the *Methods* contains 'the largest number of true and important claims' of any book in the history of ethics. In my view, Sidgwick is largely correct in his quietist non-naturalist (or perhaps broadly naturalist (Crisp 2012a)) metaethics, his intuitionist epistemology, his placing of consequentialist ethics ahead of deontology, his giving weight to both impartial and personal perspectives in the 'dualism of practical reason', and his hedonistic view of well-being. But more broadly he excels in seeing which concepts, distinctions, arguments, and positions are of most ethical significance, and in elucidating them.

The *Methods* is unarguably Sidgwick's most significant contribution to philosophy. Its publication marked an important move away from the negative tone of much work on utilitarianism in the period, written in response to J. S. Mill's widely read essay *Utilitarianism*, published in book form in 1863 (of course, Mill had his defenders). The last significant edition of the *Methods* was the seventh (1907), and, because that is now the standard edition, it will be my primary text throughout this book. But it should be remembered that Sidgwick also wrote several important essays in ethics and a classic history of ethics. Indeed, Sidgwick's published work as a whole covers many of the most significant issues in philosophy, as well as politics, economics,

religion, and even parapsychology, and one should always be prepared to read the *Methods* as merely one element in Sidgwick's overall world view. Further thought, sometimes in writing these other pieces, often led Sidgwick to make important changes to the *Methods*, and they anyway cover many issues discussed in the *Methods*. I shall not engage with this work directly, but will frequently do so indirectly in the hope of clarifying the argument of the *Methods*.

When Sidgwick died in 1900, he was widely seen as among the leading moral philosophers of his day.¹ But his reputation was eclipsed by that of G. E. Moore, and the attention paid to his work decreased. That work was nevertheless highly influential on some of the most significant moral philosophers of the twentieth century, including, in addition to Moore himself, W. D. Ross, William Frankena, John Rawls, and of course Parfit.² The shape and nature of contemporary philosophical ethics owe a huge amount to Sidgwick. In more recent years, awareness of and interest in Sidgwick have increased.³ But I believe the *Methods* receives nowhere near the attention it deserves, and one of my aims in this book is to demonstrate, through close attention to Sidgwick's arguments, how much there is to learn from him.

Much of what I say will be exegetical; and much will be critical. But this is only to treat Sidgwick as generations have treated Plato, Aristotle, Hume, Kant, and other great moral philosophers. Though there is much to gain from Sidgwick directly, we can also make progress in ethics through studying his arguments even when we believe them to be in some way misguided or mistaken.

Sidgwick himself is reported to have said that he set out to write a book in which the first word would be 'ethics', and the last 'failure', that failure consisting in his inability to reconcile egoism and hedonistic utilitarianism (Hayward 1901a: xxi n).⁴ I shall argue that, had Sidgwick allowed a greater

1. Seth 1901: 172; Sorley 1901: 168. Note the remarks of Myers (1904: 108): 'It has been said of Sidgwick that "although he was the most influential man in Cambridge, he founded no school." Not at Cambridge only, but over all the civilised world, I think that there are many of us who will say that he did found a school, and that we are his scholars.' For a different view, see Deigh 2007; and for a response to Deigh, Skelton 2010b.

2. See Skelton 2010b: 72–7; de Lazari-Radek and Singer 2014: 12–13. Skelton notes that in Ross we find almost as many citations of Sidgwick as of Aristotle, Kant, Moore, and Broad. But there was a decline: Deigh 2010: 88. In 1963, T.Y. Mullins could say that he was '[o]ne of the least popular of the nineteenth-century writers in the field of ethics' (584).

3. See e.g. Schneewind 1977; Schultz 1992b; 2004; Nakano-Okuno 2011; Phillips 2011; de Lazari-Radek and Singer 2014.

4. The page number is in fact misprinted as a second 'xix'.

role in his ethics for the idea that agents must exercise judgement in individual cases, that particular failure could have been avoided. But another failure would have been all too clear: his attempt to dismiss a reflective form of pluralistic deontology of the kind defended by Aristotle (1894) and developed by Ross (1930; 1939). The most serious objections Sidgwick makes to such a view can be avoided through reference to practical judgement, and this would leave Sidgwick's hedonistic utilitarianism depending on the positive arguments Sidgwick provides for it. As I have said, I find Sidgwick's position plausible. But his own account of the epistemic implications of disagreement requires me, and indeed other moral theorists, to suspend judgement on the question of which ethical view is correct. In other words, though something like Sidgwick's position *seems* to me plausible, I accept that I am no more justified in my belief than a non-hedonistic deontologist is in hers. This is another failure of practical reason, one much harder to resolve than Sidgwick's own.

I shall suggest also that, despite Sidgwick's recognition of the importance of clarifying ethical concepts, he was insufficiently parsimonious in his use of them. As Bernard Williams points out (1986: 19), the Socratic question is best understood as about our reasons to live—in particular, I would add, our reasons to act. The question itself does not use the notions of morality, of right or wrong, ought (except as equivalent to 'should'), duty, requirement, permission, demand, and so on. Had Sidgwick seen that his own answer to the Socratic question did not require him to use these concepts to anywhere near the extent to which he does use them, when stating his own positive views he could have avoided both having to clarify them and the various problems that arise from his own conceptions of them. In particular, he might have distanced himself sufficiently from the conceptual scheme of the morality of common sense to see that any commitment to the idea that ethics is especially concerned with the voluntary was not only unnecessary, but unwise.

Not all of Sidgwick's main claims, then, are true, and several, I shall argue, are not required for his central project, in so far as they incorporate what, for him, are superfluous concepts. I shall claim in general that Sidgwick's pursuit of this central project could have been more streamlined. He could, for example, have left aside most of the long discussion of common-sense morality in book 3. That discussion is by some margin the best account of deontological ethics we have, so it is certainly of great value and should have been published in some form. But Sidgwick's main criticisms of

common-sense morality in the *Methods* could have been stated considerably more briefly, and that, with other editorial changes, could well have led to his book's having been more widely read and studied in the earlier part of the twentieth century. That might even have significantly decreased the amount of philosophical attention paid during that period to metaethical questions the value of which in retrospect appears at least questionable (and would doubtless have appeared so to Sidgwick).

This book will be, as far as I know, the first comprehensive study of the *Methods*, and one hope I have is to bring out some of the riches that have lain largely hidden in some of Sidgwick's less widely known discussions, such as those of the virtues in book 3. The *Methods* is divided into four books. The first is 'introductory', the second concerns hedonism, the third the morality of common sense, and the fourth utilitarianism. The structure of the *Methods* is somewhat complicated, and several important topics are discussed several times, often in different books. For that reason, my approach will be thematic and I have included a list of the passages in the *Methods* most relevant to each of my chapters. I shall begin by focusing on Sidgwick's views on the nature of ethics, and the shape and content of his project. Chapter 2 will cover Sidgwick's discussion of free will in 1.5. I shall then turn in chapter 3 to Sidgwick's views on well-being and hedonism, and in chapter 4 to his intuitionist epistemology. The fifth and sixth chapters are mainly concerned with Sidgwick's own examination of the morality of common sense. I have included a chapter on virtue in particular, partly because there is much in Sidgwick to interest modern 'virtue ethicists'. Finally, in chapter 7 I shall consider egoism, utilitarianism, and the dualism of practical reason.

Sidgwick is frequently thought of as a boring writer. Certainly his style can be somewhat 'heavy and involved', as Broad puts it (1930: 143). But in fact reading Sidgwick carefully is often tremendously exciting—as exciting as coming across an unknown work on ethics by Plato or Hume. I hope that my book will encourage some to read Sidgwick and experience this excitement for themselves. The *Methods* is a philosophical gold mine, and we are only at the beginning of uncovering its treasures.

Acknowledgements

Derek Parfit, in a graduate seminar in Oxford three decades ago, first introduced me to the work of Henry Sidgwick. I am deeply grateful to him for his encouragement and advice throughout my work on Sidgwick, and in particular for many detailed and helpful comments on and discussions of drafts of this book. A first draft of the book was written in 2010–11 during my tenure of the Findlay Visiting Professorship at Boston University. I am grateful to the department for its intellectual, social, and practical support, and for discussions of Sidgwick in particular to Aaron Garrett, Charles Griswold, David Lyons, David Roochnik, Amelie Rorty, Daniel Star, and the group of outstanding graduate students who attended my weekly seminar on the *Methods*. The work was assisted by two terms of sabbatical leave from the University of Oxford and St Anne's College, and I thank these institutions and my colleagues for granting me this leave. In December 2010, Nir Eyal invited me to present a draft of chapter 3 to the New England Consequentialism Workshop. I thank him for that, and all participants for their comments. I am grateful also for the continuing and engaged support of the Uehiro Foundation on Ethics and Education. I am indebted to three Oxford University Press readers who gave me many extremely insightful suggestions. One remained anonymous; the others were Bart Schultz and Rob Shaver. For comments on all or part of the text, and/or for discussion, I am grateful to Robert Audi, Stephen W. Ball, Paul Bloomfield, Placido Bucolo, Krister Bykvist, Katarzyna de Lazari-Radek, Julia Driver, Ben Eggleston, Hortense Geninet, James Griffin, Brad Hooker, Thomas Hurka, Terence Irwin, Dale Jamieson, Leonard Katz, Paul Kelly, Nakul Krishna, Dale Miller, Kazunobu Narita, Martha Nussbaum, Justin Oakley, Mariko Nakano-Okuno, Catherine Paxton, Ingmar Persson, David Phillips, Daniel Robinson, Alan Ryan, Julian Savulescu, Philip Schofield, Sandra L. Shapshay, Peter Singer, Anthony Skelton, John Skorupski, Hillel Steiner, Philip Stratton-Lake, David Wiggins, and David Weinstein. For his encouragement and support and the continuing enlightenment provided by his groundbreaking *Sidgwick's Ethics and Victorian Moral Philosophy*, I wish to thank Jerome Schneewind. I thank also Peter Momtchiloff at Oxford

University Press for his enthusiasm and sensible advice throughout, and for assistance with production of the text Eleanor Collins, Jan Davies, Nathan Fisher, and Kim Richardson. Finally, I wish to express deep gratitude to Dennis Riddiford, a true Sidgwickian, for my first discussions of academic philosophy nearly forty years ago, and for his constant inspiration and friendship.

Earlier drafts of much of chapters 1 and 5 were published as ‘Methods, Methodology, and Moral Judgement: Sidgwick on the Nature of Ethics’, *Revue Internationale de Philosophie* 67 (2013): 397–419, and ‘Sidgwick on Virtue’, *Revue Internationale de Philosophie* 68 (2014): 21–36. I thank Bart Schultz and Justin Oakley for commissioning these pieces, and the editor for permission to draw on them. I have adapted certain passages from ‘Sidgwick and Late-Nineteenth-Century Utilitarianism’, in B. Eggleston and D. Miller (eds), *Cambridge Companion to Utilitarianism* (Cambridge: Cambridge University Press, 2014): 81–101. I am grateful to the editors and to Cambridge University Press for permission.

Contents

<i>Note on References</i>	xvii
<i>List of Key Passages</i>	xix
<i>Summary by Chapter</i>	xxi
1. The Nature of Ethics	1
1 Sidgwick's Project	1
2 The Autonomy of Ethics	5
2.1 Cognitivism and Non-naturalist Quietism	5
2.2 Is and Ought	7
2.3 Ought and Rationality	11
3 Moral Motivation	15
4 Principles and Methods	20
5 Sidgwick's Three Methods	23
6 Ethics and Politics	28
2. Free Will	32
1 The Aim and Structure of 1.5	32
2 Disinterested Action, Rational Action, and Kant	33
3 The Will	38
3.1 Voluntariness, Volition, Intention, and Responsibility	38
3.2 The Scope of the Will	42
4 Determinism and Libertarianism	45
4.1 Determinism	45
4.2 Libertarianism	48
5 The Significance of the Debate	49
3. Hedonism and the Ultimate Good	57
1 Goodness, Rightness, and the Ultimate Good	57
2 Pleasure	63

2.1	The Volitional View	63
2.2	The Feeling–Tone View	66
2.3	The Apprehension View	68
2.4	The Desirable Consciousness View	69
3	Hedonism	70
3.1	Psychological Hedonism	70
3.2	Against Synonymic Hedonism	75
3.3	Welfare Hedonism	76
3.4	Empirical Hedonism	85
3.5	Objective and Deductive Hedonisms	90
4	Objections to Empirical Hedonism	93
4.1	Practical Objections	93
4.2	Theoretical Objections	94
4.	Intuitionism	97
1	Intuition and Intuitions	97
2	Intuitionisms	99
2.1	Perceptual Intuitionism	99
2.2	Dogmatic Intuitionism	101
2.3	Philosophical Intuitionism	103
3	Self-evidence	107
3.1	Conditions for Certainty	107
3.2	Sham-axioms	112
3.3	The Self-evident Axioms	115
5.	Virtue	127
1	Intention and Motive	127
2	Virtue, Duty, and Supererogation	132
2.1	Objective and Subjective Rightness	133
2.2	The Relation between Virtue and Duty	135
2.3	Supererogation	136
3	The Nature of Virtue	138
3.1	Action and Emotion	138
3.2	Virtue and Motivation	140
3.3	The Cultivation and Value of Virtue	142
6.	The Virtues	145
1	Wisdom and Self-control	145
2	Benevolence	152

2.1	Feeling, Object, and Scope	152
2.2	Special Relations	155
3	Justice	161
3.1	Justice, Law, and Distribution	161
3.2	Contract and Custom	162
3.3	Natural Rights and Freedom	164
3.4	Desert	166
4	Laws and Promises	171
4.1	Political Obligation	171
4.2	Promises	174
4.3	Utilitarianism and Justice	176
5	Veracity	177
6	Malevolence and Liberality	181
7	Self-regarding Virtues	184
8	Courage and Humility	186
9	The Failure of Common-sense Morality	188
7.	Egoism, Utilitarianism, and the Dualism of Practical Reason	195
1	Egoism	195
1.1	A Dubious Guidance?	195
1.2	Happiness and Duty	199
2	Utilitarianism	201
2.1	Sidgwick's Form of Utilitarianism	201
2.2	The Arguments for Utilitarianism	208
2.3	Utilitarianism and Common-sense Morality	215
2.4	The Method of Utilitarianism	220
3	The Dualism of Practical Reason	227
	<i>Bibliography</i>	235
	<i>Index</i>	249

Note on References

This book is primarily concerned with the seventh edition of Henry Sidgwick's *The Methods of Ethics* (1907; all editions are London: Macmillan), and references to this are made by page number alone unless the context makes clear otherwise (some references are by book and chapter; by book, chapter, and section; and occasionally also by paragraph). It is regrettable that referencing by line numbers is in practice impossible, as is the case with many central works of modern philosophy. There is a great need for a complete scholarly edition of Sidgwick's works which would allow such referencing.

Editions of the *Methods* are referred to as M1, M2, and so on. The Preface to any edition is referred to as P along with the relevant number. So P1 refers to the preface to the first edition. Reference is then given by the page numbers as printed in the Hackett (1981, Indianapolis) reprint of the seventh edition (which adds two further preliminary pages to the original edition through including a Foreword by John Rawls). So P6 xix refers to the Preface to the sixth edition on page xix of the Hackett seventh edition. The Appendix is referred to as App.

Other abbreviations of works by Sidgwick are as follows. Page references in the text are to reprints as stated.

- AES: 'The Aims of an Ethical Society', delivered to the London Ethical Society on 23 April 1893; repr. in Sidgwick 1998: 15–30.
- APE: 'Appendix to the Preceding Essay', in *Lectures on the Philosophy of Kant and Other Philosophical Lectures and Essays*, ed. J. Ward, 1905. London: Macmillan, 461–7; repr. in Sidgwick 2000 as 'Further on the Criteria of Truth and Error': 166–70.
- CCP: 'A Criticism of the Critical Philosophy II', *Mind* O.S. 8 (1883): 313–37.
- DIO: 'Is the Distinction between "Is" and "Ought" Ultimate and Irreducible?', *Proceedings of the Aristotelian Society* 2 (1891–2): 88–92; repr. as 'The Distinction between "Is" and "Ought"' in Sidgwick 2000: 59–62.
- EEFP: 'The Establishment of Ethical First Principles', *Mind* O.S. 4 (1879): 106–11; repr. in Sidgwick 2000: 29–34.
- EP: *The Elements of Politics*, 2nd edn, 1897. London: Macmillan.

- ERRL: *Essays, Reviews, Reports, and Letters to the Editor*, in *The Complete Works and Select Correspondence of Henry Sidgwick*, online edn, ed. B. Schultz, 2nd edn, 1999. Charlottesville: InteLex Corp.
- FTDA: ‘The Feeling-tone of Desire and Aversion’, *Mind* N.S. 1 (1892): 94–101; repr. in Sidgwick 2000: 99–106.
- GSM: *Lectures on the Ethics of T. H. Green, H. Spencer, and J. Martineau*, ed. E. Constance Jones, 1902. London: Macmillan.
- GU: ‘Review of J. Grote, *An Examination of the Utilitarian Philosophy*’, *Cambridge University Reporter*, 8 Feb. 1871; repr. in Sidgwick 2000: 173–5.
- LPK: *Lectures on the Philosophy of Kant and Other Philosophical Lectures*, ed. J. Ward, 1905. London: Macmillan.
- MB: ‘Mr Barratt on “The Suppression of Egoism”’, *Mind* O.S. 2 (1877): 411–12; repr. in Sidgwick 2000: 27–8.
- NET: ‘The Nature of the Evidence for Theism’, read to the Synthetic Society, 25 Feb. 1898; in Sidgwick and Sidgwick 1906: 600–8.
- OHE: *Outlines of the History of Ethics for English Readers*, 5th edn, repr. 1910. London: Macmillan.
- PC: ‘Professor Calderwood on Intuitionism in Morals’, *Mind* O.S. 1 (1876): 563–6; repr. in Sidgwick 2000: 23–6.
- PM: ‘Public Morality’, read on 26 Jan. 1897, at a meeting of a Cambridge essay club called ‘The Eranus’; repr. in Sidgwick 1998: 31–46.
- PSR: *Philosophy, its Scope and Relations: An Introductory Course of Lectures*, 1902. London: Macmillan.
- RES: ‘The Relation of Ethics to Sociology’, *International Journal of Ethics* 10 (1899): 1–21; repr. in Sidgwick 1904: 249–69.
- SFEC: ‘Some Fundamental Ethical Controversies’, *Mind* O.S. 14 (1889): 473–87; repr. in Sidgwick 2000: 35–46.
- VOB: ‘The Verification of Beliefs’, *Contemporary Review* 17 (1871): 582–90; repr. in Sidgwick 2000: 121–8.

References to ‘CD’ are to *The Cosmos of Duty*.

List of Key Passages

The following is a list of the central passages in Sidgwick's text most relevant to each chapter of this book. References are given by book, chapter, and (sometimes) section. The list is of course non-exhaustive and readers are encouraged to use the excellent Contents and Index in the *Methods*, as well as the search facility in online texts such as that available via the University of Toronto at: <<https://archive.org/details/methodsofethicsoosidguoft>>

Ch. 1: P6; 1.1-3; 1.6.

Ch. 2: 1.5; App.

Ch. 3: 1.4; 1.7.1-2; 1.9; 2.1-4; 2.6; 3.14.

Ch. 4: 1.8; 3.1.1; 3.1.3; 3.1.4-5; 3.11.1-2; 3.13.

Ch. 5: 3.1.2-3; 3.2; 3.12.

Ch. 6: 3.3-10; 3.11.3-9.

Ch. 7: P6; 1.7.1; 2.5; 3.1.2; 3.11.7; 3.13.3-5; 3.14.4; 4.1-5; Concluding Chapter.

Summary by Chapter

1 The Nature of Ethics

This chapter covers some of the major background issues in Sidgwick's *Methods*. I begin with Sidgwick's own historical account of his development, including his search for a Kantian intuition to underpin utilitarian ethics, and the influence on him of Butler that led to the dualism of practical reason. I suggest that from the very start of the *Methods*, Sidgwick appears to be demanding a perhaps unreasonably high level of precision from ethical theory, as well as claiming a degree of impartiality in his approach which he finds it hard to live up to. I then examine Sidgwick's metaethics, explaining his cognitivism and his metaethically modest non-naturalist realism. Having set out Sidgwick's arguments against several naturalist positions, based on the ideas of God's will, conformity to nature, and self-realization, each of which Sidgwick believes fails to recognize the distinction between 'is' and 'ought', I turn to Sidgwick's views on 'ought' and his response to various non-rationalist positions influenced by Hume, centring on an appeal to the objective phenomenology of moral judgement and the claim that normativity is irreducible to any non-normative notion. Hume's metaethics depends on his account of motivation, and I explain how Sidgwick is best seen as a weak internalist about motivation, who allows for failure to be motivated by grasp of a normative reason in cases such as weakness of will or moral wrongdoing. I also claim that Sidgwick was insufficiently parsimonious in his use of ethical concepts, and could have avoided several meta-ethical debates entirely by not himself using the notion of 'ought' except in discussing the views of others. Sidgwick's own methodology depends on a distinction between methods and ultimate principles. I seek to elucidate that distinction, and argue that he would have been wiser to focus on principles rather than methods. I spell out why he chose his three methods, and claim that, though there are some difficulties with his arguments, he was nevertheless right to see ethics as a debate between consequentialism, deontology, and egoism. The chapter ends with a section on Sidgwick's views of politics, where I argue that there is a stronger case than Sidgwick allows for seeing political theory as a branch of ethics.

2 Free Will

The main aim of this chapter is to elucidate Sidgwick's discussion of free will and determinism, and to bring out some of the utilitarian or consequentialist assumptions Sidgwick appears to have been working with. Having explained the structure of 1.5, I turn to Sidgwick's preliminary discussions of disinterested action and rational action, neither of which does he see as central to understanding the notion of freedom. I explain how Sidgwick offers a form of non-standard compatibilism, according to which what matters in deliberation is not how it was caused but whether the agent can be sure how she is going to act. I raise some problems for Sidgwick's criticism of Kant, and then consider Sidgwick's philosophy of action and in particular his conception of the will. I suggest that his restriction of ethics to the voluntary was unwise, and that his own account of voluntary action is best seen not as a development of the common-sense view but as his own view, based on a cognitivist theory of intention. I then examine Sidgwick's views on the scope of the will, before turning to his discussion of the debate between determinism and libertarianism. Not all of the arguments for determinism Sidgwick finds plausible are strong, but he is right to see force in the argument from natural causation. He is also right to recognize the power of the libertarian appeal to the phenomenology of freedom. But Sidgwick's view on the unimportance of the free will debate appears to rest on a consequentialist understanding of what really matters (that is, the outcomes of actions) and a consequentialist reinterpretation of moral responsibility, involving concepts at the heart of common-sense morality, such as desert. Ultimately, it seems that Sidgwick's conclusion is that the free will debate *may* make no practical difference, if one has the capacity to substitute consequentialist conceptions of these central concepts for traditional libertarian conceptions. And even here one might wonder whether Sidgwick is right to claim that such substitutions will leave things as they are.

3 Hedonism and the Ultimate Good

This chapter begins by considering Sidgwick's position on rightness and goodness. I claim that, though there is a strong case for distinguishing reasons from values, the contrast Sidgwick draws between ancient 'attractive' and modern 'imperative' views is exaggerated. I then outline Sidgwick's view of what is good for a person in terms of what she has reason to desire, and show how this leaves him open to a 'wrong kind of reasons' objection. I turn next to various apparently different accounts of pleasure that seem to be implicit in Sidgwick's thought: the volitional view, the feeling-tone view, the apprehension view, and the desirable consciousness view. I claim that the last is best understood as equivalent to the feeling-tone view, and as Sidgwick's true view of pleasure. Having explained Sidgwick's arguments against

psychological and synonymic hedonism, I consider his own welfare hedonism, focusing on Sidgwick's objections to non-hedonistic accounts, according to which well-being consists in goods such as virtue, physical states, or non-hedonic mental states. These objections are powerful, but not so powerful as to debunk the non-hedonist position: Sidgwick's own epistemology requires him to suspend judgement on the nature of the good. The next section discusses Sidgwick's appraisal of the difficulties of measuring pleasure and pain, and his objections to 'objective' and 'deductive' hedonism. I end with some discussion of what Sidgwick said, and might have said, in response to various practical and theoretical objections to hedonism.

4 Intuitionism

This chapter concerns Sidgwick's ethical epistemology. I begin by explaining Sidgwick's cognitivist view of intuitions as apparently self-evident beliefs. I express some doubt about Sidgwick's quick dismissal of 'perceptual' intuitionism before moving on to his objections to 'dogmatic' intuitionism and his arguments for 'philosophical' intuitionism. Here I claim that Sidgwick should have attached less weight to the morality of common sense than he did. I then move on to the notion of self-evidence, and Sidgwick's conditions for 'highest certainty'. I claim that Sidgwick failed to follow through properly on the implications of his non-dissensus condition for his own ethics. After outlining Sidgwick's critique of what he calls 'sham-axioms', I discuss the principles of justice, prudence, and benevolence which Sidgwick himself found self-evident, suggesting again that dissensus poses difficulties, though not insurmountable ones, for Sidgwick.

5 Virtue

In this chapter I focus on Sidgwick's views on virtue in general, as opposed to those on particular virtues, which will be discussed in the following chapter. I begin with an examination of Sidgwick's views of the ethical significance of intention and motive, arguing that his giving too much credence to the notion that ethics is concerned with the voluntary again is in tension with his own utilitarian approach. I then consider Sidgwick's distinction between 'objective' and 'subjective' rightness, and the best way to understand that distinction in relation to virtue. Having looked at Sidgwick's important discussion of the difference between virtues and talents, I find some problems in Sidgwick's position on supererogation. The third section of the chapter concerns the nature of virtue itself. I argue against the view that we should see being virtuous primarily as a property of a person rather than her acts, and then consider Sidgwick's position on moral motivation and the sense of duty.

The chapter ends with a discussion of Sidgwick's views on the cultivation of virtue, and I raise the question whether he is right to doubt that virtue, as a trait of character, could be valuable in itself.

6 The Virtues

This chapter consists primarily in discussions of Sidgwick's treatments of individual virtues. I argue that Sidgwick distorts practical wisdom into a form of self-control, partly because of his antipathy to the notion of judgement, but also because of his restriction of ethics to the sphere of the voluntary. I suggest that Sidgwick's account of benevolence rests too much on his own hedonistic account of the good, and then examine his fourfold distinction of the duty into involuntary duties, voluntary duties, duties of gratitude, and duties of pity. I claim that once again Sidgwick's own utilitarian views appeared to be affecting his own interpretation of common-sense morality. I then turn to Sidgwick's discussion of justice, and in particular its relation to law, contract, and custom. I show how Sidgwick's emphasis on the importance of not disappointing expectations is worth greater attention than it has received in the literature, and how his criticisms of natural rights theories and desert theories anticipated various later arguments. Sidgwick himself offers a separate discussion of laws and promises, and I begin by looking at his account of political obligation and his apt criticisms of consent theories. I try to bring out how much insight Sidgwick shows in his account of the limits on the duty to keep a promise. After explaining Sidgwick's view on the role of justice within a utilitarian theory, I discuss various other virtues: veracity, non-malevolence, liberality, the self-regarding virtues, courage, and humility. I then focus on Sidgwick's overall analysis of the failure of common-sense morality, bringing out how Sidgwick often confronts common-sense moral principles with a 'clarity-non-dissensus dilemma'. I conclude that, though there is a huge amount to learn from Sidgwick's examination of common sense, his criticisms are flawed by a failure to recognize the role that a capacity for individual judgement can play within an ethical theory.

7 Egoism, Utilitarianism, and the Dualism of Practical Reason

In this chapter I examine Sidgwick's views on egoism, utilitarianism, and the conflict between the two that he called 'the dualism of practical reason'. I show that, though Sidgwick finds egoism somewhat objectionable, he sees an intuitive grounding for it in the metaphysical distinction between one individual and another. I suggest that the view can be made consistent with a Humean view of personal identity, and offer some defence of Sidgwick's general view on the

plausibility of egoism. I also find persuasive Sidgwick's view that there is a potential conflict between happiness and duty, as ordinarily understood. Continuing on to utilitarianism, I argue that Sidgwick's 'objective' version of the view is as plausible as more recent 'subjective' or probability-based versions. I then examine Sidgwick's views on the moral status of non-humans and future generations, and raise some questions about what appears to be the advocacy by Sidgwick of a restricted form of egalitarianism. Having outlined Sidgwick's intuitionist argument for utilitarianism, I then turn to his 'proof' of the view in book 4, which attempts to persuade common-sense moralists that utilitarianism can be seen as 'controlling and completing' common-sense moral principles. This proof, I claim, might succeed in certain cases, though it is open to a defender of common-sense morality to claim that her principle of benevolence can do at least much of the work done by the utilitarian principle in Sidgwick's account. Sidgwick recognizes that he should take a view on what kind of value 'moral value' is, and I applaud his account. I then consider Sidgwick's careful and balanced discussion of the relation between utilitarianism and common-sense morality, including his famous admission that utilitarianism may recommend its own partial suppression. The final section discusses Sidgwick's dualism of the practical reason. Through understanding Sidgwick's versions of egoism and utilitarianism as making consistent yet universal demands, I show why Sidgwick appeals to the possibilities that sympathy may play an important role in happiness or that the world is morally governed by God as resolving the dualism of practical reason. Neither of these strategies works, but I suggest that had Sidgwick seen his principles of prudence and benevolence as pro tanto principles guiding judgement about particular cases, the dualism could have been avoided.

I

The Nature of Ethics

I Sidgwick's Project

Henry Sidgwick was thirty-six when *The Methods of Ethics* was first published in 1874. For well over a decade he had been planning a book that might reconcile 'moral sense', or intuition, with utilitarianism, and he revised the book in various editions for the rest of his life (Sidgwick and Sidgwick 1906: 75). Sidgwick died before completing the sixth edition, which was seen through the press by E. E. Constance Jones. In the preface Jones included some illuminating notes for a lecture by Sidgwick on the development of his ethical views. Although Sidgwick is often, with some justification, described as the third of the great 'classical utilitarians', we can see from his own account of his intellectual journey that, though it began with utilitarianism strictly understood, it soon departed from there and was never to return.

Sidgwick tells us that he was initially committed to Millian utilitarianism, which he found liberating, in contrast to the 'arbitrary pressure' of the dubious, confused, or dogmatic moral rules he had been taught. That commitment sat alongside antipathy to the views of the influential Cambridge philosopher William Whewell, whose *Elements of Morality* (1845) led Sidgwick to the view that 'intuitional' ethics was, in comparison with mathematics, 'hopelessly loose'.¹ Sidgwick understood Mill to hold that each person seeks her own happiness (psychological hedonism) and that each person ought to promote the happiness of all (ethical hedonism). He found both views attractive, not yet seeing the potential inconsistency between

1. 'Intuitional ethics' is what Sidgwick elsewhere calls 'intuitionism': the normative, non-utilitarian view developed out of reflection on common-sense morality. Somewhat confusingly, Sidgwick uses the same name for the epistemological view that certain moral principles are self-evident and can be known through a faculty of intuition (see Broad 1930: 206; CD 4.2).

them. As he came to recognize the possibility of conflict between ‘interest’ and ‘duty’, Sidgwick began to think seriously about egoism as a normative view, and concluded that it is the opposition between ‘interest’ (that is, self-interest) and duty or the general good that is central to ethics, rather than that between ‘intuitions’ and hedonism. This, he says, explains the structure of the *Methods*, in which, after an introductory book, egoism (interest) is discussed in book 2, intuitionism (duty) in book 3, and utilitarianism (the general good) in book 4. Further, Sidgwick decided that, despite his earlier aversion to intuitionist ethics, the only way to ground a utilitarian justification for sacrificing one’s own happiness for the sake of others was through a fundamental ethical intuition: ‘I must somehow *see* that it was right for me to sacrifice my happiness for the good of the whole of which I am a part’ (P6 xviii).

Where was he to find such an intuition? Sidgwick returned to Kant, and, though impressed by Kant’s view that what is right for any person must be right for all persons in similar circumstances, felt that this notion of universalizability was insufficiently substantive to ground a principle of duty in opposition to egoism. This led Sidgwick to Butler, who he believed also accepted a ‘dualism of the practical reason’. It was under the influence of Butler that Sidgwick rejected psychological hedonism and accepted the existence of disinterested, other-regarding motivation. Further, since Butler’s powerful critique of utilitarianism also worried Sidgwick, and Sidgwick realized that he himself was already an ‘intuitionist’ by accepting both the Kantian thesis about rightness and the need for an intuitionist foundation for utilitarianism, he decided to reconsider intuitionist ethics itself. Unsurprisingly, Sidgwick decided to study not Whewell but Aristotle, seeing his *Nicomachean Ethics* as an impartial attempt to make consistent the common-sense morality of his day. Sidgwick set out to do the same for nineteenth-century British morality, and 3.1–9 was the first component of the *Methods* to be completed. The upshot was that, though he failed to find any further intuitions he could accept, he came to see common-sense morality as a system itself based on the promotion of general happiness, to be revised in the light of general happiness only in exceptional cases. Further, despite having failed to find any way to resolve the conflict between egoism and utilitarianism, Sidgwick decided to publish his book, finding it a ‘great comfort to have got it out’ (Sidgwick and Sidgwick 1906: 295).

Sidgwick’s aim, then, was to ascertain which—if either—of egoism and utilitarianism was correct, by reflection on the possibility of securely

grounding either on a secure intuitionist basis. But he wished also to inquire into whether elements of traditional intuitional or non-utilitarian ethics could be so grounded. Sidgwick does not say how he would have dealt with such non-utilitarian principles had they appeared justified, though we might presume that, just as in the case of egoism and utilitarianism, he would have sought to make them consistent with one another.² Philosophical ethics, Sidgwick tells us, aims like science to be ‘systematic and precise’ (1; see Irwin 2009: 3: 479–84).³ He later says that the assumption that moral rules should be precise ‘naturally belongs to the ordinary or jurial view of Ethics as concerned with a moral code’ (228), and provides an argument for this view based on an apt analogy with law. If a law were vague, we would think it to that extent unreasonable: anyone subject to a legal obligation ought to be in a position to know what it is. Similarly, a moral philosophy which left it unclear on some occasion exactly what a person’s obligations were would, to that extent, have failed.⁴

A good deal of law is indeed highly precise. The UK Representation of the People Act 1969, for example, leaves no doubt about when a person becomes eligible to vote in a parliamentary election: on their eighteenth birthday. But some law is less precise. Consider, for example, the definition of obscenity in the Obscene Publications Act 1959, still in force in the UK:

For the purposes of this Act an article shall be deemed to be obscene if its effect or (where the article comprises two or more distinct items) the effect of any one of its items is, if taken as a whole, such as to tend to deprave and corrupt persons who are likely, having regard to all relevant circumstances, to read, see or hear the matter contained or embodied in it.

2. Sidgwick would have agreed with Mill in the *System of Logic*: ‘There must be some standard by which to determine the goodness or badness, absolute and comparative, of ends, or objects of desire. And whatever that standard is, there can be but one: for if there were several ultimate principles of conduct, the same conduct might be approved by one of those principles and condemned by another; and there would be needed some more general principle, as umpire between them’ (Mill 1961–91: 8.951; see 1998: 1.3.23–35).
3. Note the epigraph to the *Methods* from Plato’s *Republic*, which advocates precision and clarity about the most important matters. Cremaschi (2008: 176) draws an interesting link between Sidgwick’s scientism and his noticeably dry style, a style he himself fully recognized and saw as to some degree ‘repellent’ (Sidgwick and Sidgwick 1906: 295). De Lazari-Radek and Singer (2014: 15) suggest that Sidgwick may have taken the term ‘method’ itself from J. S. Mill’s discussion of scientific methods in the *System of Logic*. It is worth remembering that Whewell, to whom Sidgwick was so opposed, saw philosophical ethics as analogous to the theory of geometry (1845: vi–ix).
4. Singer (1986: 69) rightly suggests that 102 implies that Sidgwick thought an ethical view should not leave any ‘practical questions unanswered’. See Schultz 1975: 348.

If I have written some potentially obscene article, and am considering its publication in the UK, the law will not tell me whether or not it is safe for me to publish it. I have to rely on my judgement about the likely effects of its publication, and whether they might be described as depravation or corruption. Now it might be said that Sidgwick is right that, to this extent, the law is a failure, and legislators should seek further precision. But even if they did—perhaps by spelling out further what is meant by depravation and giving examples of what is and what is not to count as obscene—there will be an ineliminable role for judgement on the part of citizens. A law is not a failure if it is *reasonably* clear, and relies only to a *reasonable* degree on individual judgement. Exactly what counts as reasonable or not in ethics is a highly important question, and I shall return to it later in this book (see CD 6.9).⁵

Given the prevalence of apparent conflict between the considerations picked out as relevant by utilitarian and non-utilitarian principles, consistency between them, if it were even possible, would have required any single ultimate principle to become hugely complicated and detailed. And at that point Sidgwick would almost certainly have rejected it as intuitively insecure (see e.g. 342–3), and as presenting him with yet another failure. To avoid such complication would have required him to accept the possibility of decision-making based on the weighing of a plurality of ultimate principles against one another in individual cases analogous to the kind of legal judgement mentioned in the previous paragraph. As we shall see (CD 7.3), had he allowed the possibility of such judgement, he could have avoided not only the danger of conflict between utilitarian and non-utilitarian principles, but the actual conflict between egoism and utilitarianism he so signally failed to resolve at the close of the *Methods*.

Also important to Sidgwick's own conception of his project is his avowed impartiality (see Irwin 2009: 439–40; 518). At P1 viii, Sidgwick claims that the protreptic goals of moral philosophers have distorted the development of ethics as an objective enterprise, and announces his intention to focus on the 'methods' of ethics themselves and not their practical results:

5. In his discussion of Aristotelian ethics (OHE 66), Sidgwick does note Aristotle's virtue of 'equity' (1894: 5.10) 'as a kind of justice superior to that which is realised by strict adhesion to the letter of law, and *rightly* overruling it' (my italics). The implication is that Sidgwick agrees with Aristotle on the need for such a virtue, since he is not merely reporting Aristotle's view (the next paragraph goes on to discuss and partly respond to a potential objection to Aristotle). On the following page, he moves on to Aristotle's account of practical wisdom, failing to mention its involving a quasi-perceptual capacity to see what is called for in particular cases (see e.g. 1894: 1142a25–30; 1143b11–14).

I have wished to put aside temporarily the urgent need which we all feel of finding and adopting the true method of determining what we ought to do; and to consider simply what conclusions will be rationally reached if we start with certain ethical premises, and with what degree of certainty and precision.

Later, at 13, Sidgwick claims that his only ‘immediate end’⁶ has been to clarify the three methods themselves and that he has therefore ‘refrained from expressly attempting any such complete and final solution of the chief ethical difficulties and controversies as would convert this exposition of various methods into the development of a harmonious system’ (see also 338). This is mysterious.⁷ Sidgwick famously considered the *Methods* a failure, but not because it left the three methods unclear. The failure consisted in the dualism of the practical reason—the conflict between egoism and utilitarianism that made impossible the development of a ‘harmonious system’.⁸

2 The Autonomy of Ethics

2.1 Cognitivism and Non-naturalist Quietism

Sidgwick’s disparaging contrast between intuitionist ethics and mathematics strongly suggests that he saw the latter as the model at which philosophical

6. Sidgwick speaks also of mere articulation of the methods as his ‘primary aim’ (14), and this might suggest that he sees practicality as a secondary or less immediate aim. But he at once goes on to claim that throughout the book he has sought to focus only on the processes and not the results of ethical thought, and has ‘never stated as [his] own any positive practical conclusions unless by way of illustration’.
7. See Singer, M. 1974: 432–3; 445. For an interpretation more in line with Sidgwick’s own account of his aims, see Schneewind 1977: 192–3. Schneewind provides an excellent explanation of why Sidgwick was motivated to carry out an impartial study of common-sense morality. But everything he says is consistent with the notion that this study led Sidgwick to an argument for utilitarianism.
8. At 13–14, Sidgwick notes that spelling out the implications of different ethical principles is likely to make some of them appear less plausible than they did initially. But he does not always, once he has explained the implications of the various methods, leave it up to the reader to draw her own conclusions. Rather, he seeks to bring out the ‘inevitable imperfections’ (361) of common-sense morality in particular through demonstrating that its principles appear plausible only if they are left without elucidation, and once spelled out are not self-evident. Sidgwick might well be understood to be bringing out imperfections in egoism and utilitarianism when he outlines serious problems with the practicality of hedonism in 2.3; and yet he does not describe them as imperfections. All this, however, is not to deny that he is impartial when considering the self-evidence of the various ultimate principles he considers. Sidgwick fails to see both that impartiality in judgement does not require one to abstain from judgement, and that he himself does not abstain from judgement. It may be that his desire to create an impartial scientific ethics led to a degree of self-deception about his real methodology.

ethics should aim. Indeed, he is happy to see philosophical ethics as scientific in so far as it aims at being systematic and exact (1–2). But calling it a science, he suggests, may be misleading, since the objects of ethics are quite different from those of the ‘positive sciences’; ethics is better described as a *study*.⁹ Sidgwick notes that psychologists and sociologists who seek to explain human action individually or collectively often begin to speak of the actions or institutions they study as ‘good’, ‘bad’, ‘right’, or ‘wrong’, and, in doing so, they move out of their ‘positive’ discipline into ethics or politics. The attempt to *explain* human action, sentiments, and judgements is ‘essentially different’ from the attempt to decide which of these is *right* or *valid*.¹⁰ Ethics is not a form of ‘naturalism’, in so far as it sees *what ought to be* as an object of knowledge (PSR 24; 76; see P1 vii), while it is usually held that a science must be an inquiry into some ‘department of actual existence’.

Sidgwick is, then, both a cognitivist—in the sense that he accepts that there can be ethical knowledge, at least in principle—as well as a non-naturalist, in so far as he denies straightforward naturalism (see Phillips 2011: 12).¹¹ Is he also committed to normative, ethical, or moral realism, if we understand such positions to involve a substantive metaphysical commitment to moral or ethical facts?¹² Sidgwick never speaks of such facts.¹³ And

9. At 4, Sidgwick is ready to speak of ethics as a science. At 99, he uses the phrase ‘scientific Ethics’, which replaces ‘a science of Ethics’ from M5 99. Sidgwick appears to have become more cautious in his usage: the first sentence of M1 describes ethics as the ‘Science of Practice or Conduct’. Nevertheless, he is prepared to use the phrase ‘the Science of Right’ from the list of aims set for itself by the London Ethical Society in speaking of the aim of the moral philosopher (AES 23).
10. By ‘valid’ here, and often elsewhere, I take Sidgwick to mean ‘justified’. That is, a valid claim is one for which there is justification available. See CD, chapter 7 n. 26.
11. G. E. Moore (1903: 17–21) credits his own non-naturalism to Sidgwick. At 1903: 17, in discussing the ‘naturalistic fallacy’, Moore claims that Sidgwick is the only writer he knows who has clearly stated that ‘good’ is indefinable; but his only reference is to 1.3, which discusses judgements of rightness or reasonableness; see especially 25. See Frankena (1956: 541), who correctly notes that Sidgwick’s view that rightness is indefinable can be found in the work of Richard Price and many others.
12. Phillips (2011: 12–14) attributes to Sidgwick what he calls the ‘realist conceptual thesis’: ‘It is part of our moral concepts that there is such a thing as moral truth and error’. His evidence comes from the preface and the first chapter of M1. As Phillips himself says, these passages ‘treat . . . realism [that is, the view that there is moral truth and error] as a presupposition of all moral enquiry’. But a presupposition does not have to be conceptual. So when Sidgwick says in the passage from the first chapter that the view that rightness depends on the agent’s beliefs about rightness implies that the ‘common notion of morality must be rejected’, he need not be taken as making a conceptual claim. What he is rejecting is not a particular understanding of the concept of morality, but a conception of it.
13. See Schneewind 1963: 141; 1977: 205; Shaver 2000: 263–4. At DIO 61 Sidgwick contrasts the ‘world of duty and the world of fact’.

his reference to the subject matter of positive science as some aspect of ‘actual existence’ suggests that he was not inclined to think in those terms.¹⁴ For him, it was enough that we can have knowledge of moral truth (34), and that philosophical ethics be seen as the discipline through which we might acquire such knowledge.¹⁵

For this reason (see Schneewind 1977: 222; Shaver 2000: 262–6; Phillips 2011: 27–32), Sidgwick appears to be in a position to sidestep the charge of metaphysical extravagance often made against ethical non-naturalism (see esp. Mackie 1977: 38–42). Of course, in some weak sense he must be committed to ethical facts, if the postulation of such facts amounts to nothing more than the claim that ethics has an object, irreducible to non-ethical facts, which we can cognize (see Phillips 2011: 27–32). But Sidgwick’s moral metaphysics is derivative from his epistemology. He does not begin with the claim that there is some ‘special realm’ of moral facts, like the Platonic forms, and then provide an epistemology in the light of that.¹⁶

2.2 Is and Ought

At 78–83, Sidgwick notes that his list of three ultimate reasons suggested for further attention—happiness, excellence, and duty—may appear to leave out some important further options for ultimate reasons: that something is God’s will; that it promotes ‘self-realization’; or that it is in conformity to

14. There is the odd exception. At M1 x, for example, Sidgwick says of moral truth itself that it ‘really exists’.
15. Shaver (2000: 273; 2006: 14) suggests that, because Sidgwick’s commitment was to rationalism rather than cognitivism, he might have been satisfied by ‘ambitious’ forms of non-cognitivism, such as those developed in the twentieth century by R. M. Hare or Allan Gibbard. But Sidgwick insists that moral judgements are ‘cognitions’, and may well not have seen as sufficiently powerful the rational requirements on such judgements found even in the most sophisticated non-cognitivist theories (see Phillips 2011: 21–32; cf. Parfit forthcoming). Phillips suggests that Sidgwick’s dualism might imply that he ought to have been, and perhaps sometimes was, an error theorist: ‘The central thought is that if we have a body of beliefs that we take to express positive truths about a certain subject matter, but these beliefs turn out to involve a fundamental contradiction, then we are not entitled to think that there are positive truths about this subject matter’ (45). I am inclined to think that Sidgwick would have considered this position both too quick and too pessimistic. But there is also an internal question. If he accepted what Phillips calls (2011: 12) the realist conceptual thesis (it is part of our moral concepts that there is moral truth and error; see CD, chapter 1 n. 12) and the realist substantive thesis (there is moral truth and error), and on the basis of these would have denied non-cognitivism (Phillips 2011: 28), then why would this not have led him to reject error theories for the same reasons?
16. Recent ethics has seen a good deal of convergence on a Sidgwickian position: see e.g. Nagel 1986: ch. 8; Scanlon 1998: ch. 1; 2014; Skorupski 1999; 2010: ch. 17; Crisp 2006: ch. 2; Parfit 2011: pt. 6; forthcoming. See Phillips 2011: 27–32.

nature. Sidgwick is ready to accept the immediate attraction of such views, since they might appear to provide a foundation for what ought to be in what actually exists. He then claims that, for this very reason, they are the concern not of ethics as he has defined it (as an autonomous discipline) but of philosophy more generally, which is concerned with the relations of all objects of knowledge.

This restriction of the scope of ethics is somewhat unsatisfactory. Sidgwick might be proposing his conception of an autonomous ethics for pragmatic reasons—as a research programme which, though it leaves out some significant options, will be for that reason more manageable. But this will leave his conclusions as merely temporary hypotheses, which, even if they appear highly plausible, will require testing against the ethical conclusions from ethics more broadly construed as part of philosophy. This seems especially regrettable given the pessimistic conclusion of the *Methods*: a proponent of one of the three alternatives might well be hopeful of greater philosophical success.

But Sidgwick's restriction is best seen not as purely heuristic, since he continues:

The introduction of these notions into Ethics is liable to bring with it a fundamental confusion between 'what is' and 'what ought to be', destructive of all clearness in ethical reasoning; and if this confusion is avoided, the strictly ethical import of such notions, when made explicit, appears always to lead us to one or other of the methods previously distinguished.

Sidgwick proceeds to examine the 'divine will' view first, since here, he says, the connection between 'what is' and 'what ought to be' is clear. He raises the question of how we might ascertain God's will, and suggests it must be through revelation, reason, or some combination of both. Revelation, he suggests, does not fall within the domain of ethics as he understands it. Given that the epistemology of revelation might turn out to be analogous in important ways to Sidgwick's own philosophical intuitionism, this is not a strong response to the revelational position. He might rather have allowed for the view, and then impaled it on one or other horn of the *Euthyphro* dilemma. If the idea is that God's willing in itself gives us a reason to act, this appears arbitrary, and also fails to explain why—on most theistic accounts—God is said to will actions for which there often seem to be excellent reasons independent of his willing (see 31). But if God is seen merely as the best guide to what is independently right, then the view becomes more

strictly ethical, and Sidgwick can run his argument that such positions can be categorized within the scope of the ethical theories he does discuss. It is a form of this latter argument that he uses against the rationalist theist, claiming that her ethical position will be either one of his original three, or a version of the 'self-realization' or 'nature' views.

Sidgwick then moves to a discussion of the conformity-to-nature view (CTN). He claims first of all that human nature must be seen as designed, if the view is to be at all plausible. If nature were just a matter of aimless change, it becomes quite unclear how it could ground ends or moral rules (later, at 83, Sidgwick applies this claim to evolutionary theory). So let us consider the design view. First, to avoid the ethical conclusion that 'anything goes', we need to distinguish the notion of natural impulses from those that humans actually experience. But we cannot do this by claiming that the supremacy of reason is natural, since the starting assumption of the view is that reason requires conformity to nature. CTN-theorists have usually understood the natural to be either what is common or usual, or what is original or earlier as distinct from what develops later. But Sidgwick sees no ground for thinking that nature is opposed to the unusual or prefers what is earlier in time. Some widely admired impulses—such as enthusiastic philanthropy—are uncommon and emerged later than several other impulses often thought less valuable. Indeed, many 'later' developments, such as certain societal institutions, seem beneficial, and it would anyway be arbitrary to see them as beyond the 'purposes' of nature. Nor can any substantive ethical conclusions be drawn from premises about the physical make-up of humanity. Reflection upon the actual social relations within any human society will take us to the question of what such relations *ought* to be, and our original triad.

Although many variants of CTN do fall to one or more of Sidgwick's objections, one appears to remain standing. Sidgwick's restriction of CTN to versions which postulate design in nature is unmotivated. Where human nature came from is irrelevant; what matters, according to CTN, is what it is. Many have accepted as plausible the Aristotelian view that the good of any being will depend on its nature, and that such a good is promoted through increasing the perfection of that nature. The question is what that nature is; and what emerges from reflection on the perfectionist tradition is that many thinkers have seen the acquisition of knowledge, social relationships, and so on as both characteristically human and aspects of the human good, independently of their relation

to consciousness, pleasure, or moral excellence.¹⁷ Here we see an alternative to Sidgwickian hedonistic happiness and moralized perfection emerging as a candidate for ultimate end, an end which might be grounded in a form of perfectionism that avoids any ‘is’/‘ought’ objection through its reliance on a self-standing normative premise linking the good with human nature (see CD 1.5).

What about self-realization? Sidgwick (80 n) postpones discussion of this until 89–91 in the following chapter, which concerns different interpretations of egoism. Sidgwick begins by noting the Hobbesian and Spinozan versions of egoism as a principle of self-realization. He suggests that the intellectualism implicit in Spinoza’s conception would seem quite implausible to someone who saw the human good as consisting in action, the creation of artistic beauty, or obedience to reason or conscience. Egoism cannot be based on an appeal to those impulses which our self finds prominent, since any such impulses could meet this criterion. But it might be said that we should exercise our faculties and so on in ‘due proportion’. There are two ways in which this proportion might be understood. It might be taken to refer to whichever combination of tendencies we were born with; so we might be urged to ‘be ourselves’. But this can be plausible only as an instrumental recommendation. If adhering to our original nature made us very unhappy, no one would recommend doing so. Alternatively, the due proportion in question can be seen as merely another term for perfection.

Sidgwick’s conclusion is that the notion of ‘self-realization’ is too vague to be of service in philosophical ethics. This comes as something of a surprise, since he himself must have thought he had gone some way to clarifying the position in these pages of 1.7. On the face of it, what Sidgwick should have concluded is that self-realization theories are really versions of moralized or non-moralized perfectionism; and he might then have referred the reader to his criticisms elsewhere of, respectively, common-sense morality and the claim that there are goods independent of pleasure.

17. Sidgwick clearly distinguishes a non-moralized form of perfectionism from hedonism at 97. It has to be admitted that a proponent of this version of the CTN view will face problems if she admits there are elements of human nature—such as, say, our propensity to unreflective aggression—which we would not wish to include as part of the human good. But of course she does not have to allow that common traits emerge from human nature properly understood; they may be deviations from it.

2.3 Ought and Rationality

Sidgwick seeks his answer about what we ought to do through examining the different methods to be found in common-sense morality. So, to that extent, he is himself moving from sociological description to normative ethics, though his justification for the sociology is itself grounded in ethical epistemology. He is careful to distinguish the ultimate ought-judgements he is interested in from what Kant calls ‘hypothetical imperatives’ (6–7). These apply only to those who have already accepted the relevant end. If my art teacher tells me I ought to hold my brush differently, she is saying this on the assumption that I want to improve my technique. If I’m in the class just to get out of the house, however, her ought-judgement no longer applies to me. Contrast that with the judgement that anyone *ought* to care about her own happiness. This is a categorical judgement; whether or not one *wants* to be happy or to care about happiness is irrelevant.

If the ultimate ends I ought to pursue are independent of my wants or desires, it is tempting to follow Kant and see them as ‘prescribed by Reason’ (23–8). If we do so, we shall be denying a Humean account of reason as merely the slave of the passions. Sidgwick’s enquiry into this question is typically empiricist: he turns to introspection, and describes his own experience—which he plausibly believes to be widely shared—of what we mean by a conflict between reason and desire, mentioning as an example the bodily desire for indulgence judged to be imprudent. His denial of the Humean position is itself couched in terms of his earlier distinction between ethics and positive science:

I hold . . . that the ordinary moral or prudential judgments which, in the case of all or most minds, have some—though often an inadequate—influence on volition, cannot legitimately be interpreted as judgments respecting the present or future existence of human feelings or any facts of the sensible world; the fundamental notion represented by the word ‘ought’ or ‘right’ which such judgments contain expressly or by implication, being essentially different from all notions representing facts of physical or psychical experience. (25)¹⁸

18. Phillips (2011: 12) calls the view stated here ‘the fundamental distinction thesis’. Moore is Phillips’s paradigm of a philosopher who claims to have a ‘master argument’ against any form of ethical naturalism. Sidgwick, in contrast, is said to offer only piecemeal, negative arguments against particular versions of naturalism (17). So ‘we should not expect to find in Sidgwick any diagnosis for [the] . . . failure [of these particular analyses] beyond the general idea that they fail to capture something central to the fundamental concept expressed by “right”, “ought”, etc.’ (18). I fail to see, however, why this view does not itself amount to a ‘master argument’. Sidgwick’s claim is that any form of naturalism will be found to contradict the fundamental

Sidgwick's argument against Hume here is somewhat problematic. Hume's position need not be understood as the view that moral judgements themselves must be seen as judgements about sensible facts. Rather, they may be seen as expressions of 'calm passions', which themselves are not based on any rational judgement (Hume 2007: 2.3.3.8).¹⁹ The Humean, then, might agree that the word 'ought' has a different sense from any set of mere assertions of facts, but deny Sidgwick's suggestion that our moral phenomenology is best understood in rationalist terms as the cognition of normative truth.

These issues of course received a great deal of attention after Sidgwick's death, and Sidgwick's discussion does anticipate several later developments. For example, in further criticizing the view that 'rightness' is always to be understood hypothetically, as an attribute of means that 'ought' to be taken to some end, Sidgwick notes that we cannot understand Bentham, despite his own suggestions to the contrary, as claiming that the word 'right' means 'conducive to the general happiness' (26 n; see 109; CD, chapter I n. 11).²⁰ For his ethical view would then be a mere tautology. Further, the criticism of the instrumentalist view of rightness is itself powerful. Sidgwick points out that just actions, for example, are often said to be right in themselves, independently of any end, and also that we often see it as right to adopt certain ends, such as the common good.

Sidgwick also poses some serious problems for those versions of Humeanism according to which moral judgements such as 'Truth ought to be spoken' are equivalent to factual judgements such as 'The idea of truth-speaking excites in me a feeling of approbation'. Sidgwick admits that such a feeling will ordinarily accompany a moral judgement, but finds the meta-ethical position itself to be 'absurd' because it cannot explain why we take 'Truth ought to be spoken' and 'Truth ought not to be spoken' as contradicting one another. The suggestion that, on reflection, all I affirm when I make a judgement such as 'the air is sweet' is 'the air pleases me' may well be true in certain cases. But it is not so with the moral emotions: 'The peculiar

distinction thesis. This appears to be structurally analogous to Moore's suggestion that any form of naturalism will close certain questions which we believe are open. So in the case of Bernard Williams, for example, we can predict that his failure will consist in unacceptably denying the fundamental distinction thesis (20–1); and in the case of Leslie Stephen, we find Sidgwick wheeling out that thesis using the very same terminology as in the key statement in the *Methods* (cited in Phillips 2011: 27).

19. More rationalist readings of Hume, focusing for example on his notion of the 'steady and general point of view' (2007: 3.3.1.15), are of course available; see e.g. Wiggins 1995: 300–3.

20. It is this note by Sidgwick to which Moore must be referring at 1903: 17 n. 1. See CD n. 13.

emotion of moral approbation is, in my experience, inseparably bound up with the conviction, implicit or explicit, that the conduct approved is “really” right—*i.e.* that it cannot, without error, be disapproved by any other mind’ (27). Sidgwick’s argument against Humeanism, then, rests on a phenomenological appeal of a kind which would undergird many defences of cognitivist positions in the century to follow. He goes on to provide an example of the distinction he has in mind between a moral sentiment against acting in some way and mere repugnance. Consider a case of someone who habitually feels a sentiment in favour of veracity and then becomes convinced that, in certain circumstances, she ought to lie. The ‘quasi-moral’ (28) feeling of repugnance she may feel when lying, as she believes she ought to do, seems quite different from the moral sentiment that led her to telling the truth as a matter of virtue earlier.²¹ A similar example enables us to dispose of the suggestion that moral judgements concern not the individual’s own moral approbation but that of her society: consider the case in which someone comes to a moral view in opposition to that generally held in her society.

Sidgwick then turns to a metaethical position which did not survive into the twentieth century: that moral judgements are claims about the pain that the agent will suffer if she fails to do what the judgement requires (29–31). Sidgwick notes—quite correctly, I believe—that there is something plausible in this position, namely, the analogy on which it rests between law and morality. In the case of positive law, we might well claim that some law is not genuinely established in a society unless its violation results, or may result, in punishment. But in the case of morality, not only will we distinguish ‘P ought to ϕ ’ from ‘P will be punished by public opinion if he fails to ϕ ’, but we will sometimes make that former claim knowing that no such punishment will occur. In general, reflection will demonstrate the distinction between duties such as those of mere politeness, which exist purely as a result of their imposition by public opinion, and strictly ethical duties. And in the same way, those who hold a divine command view of ethics will see the claim ‘P ought to ϕ ’ not as equivalent to ‘God will punish P for not- ϕ -ing’ but as providing the *grounds* for God’s punishment.

In his own positive account (33), Sidgwick distinguishes between ‘ought’ in a narrow sense, in which ‘ought’ implies ‘can’, and a broader sense in which I might think, for example, that I ought to know what a wiser person

21. Sidgwick might also have mentioned the phenomenology of the agent’s decision to lie based on the judgement that, in this case, it is right.

would know, even though I am quite aware that I am unable to will myself into that position here and now. In this latter sense, ‘ought’ implies some ideal which I ought, in the narrow sense, to approximate as far as possible. But Sidgwick goes no further than this. Indeed, he suggests that any reductive account of ‘ought’ will fail, since the notion underlying it and other terms such as ‘right’ is ‘too elementary to admit of any formal definition’ (32).²² This is not to deny that we might be able to explain how that notion emerged. But its content is not itself to be understood in terms of that explanation: ‘I know of no justification for this transference of the conceptions of chemistry to psychology.’ All we can do to elucidate ‘ought’ is to explain its relation to certain other notions.²³

Let me end this section by returning to the distinction between categorical and hypothetical imperatives. Initially, Sidgwick appears to wish to avoid confrontation with Humeans: at 8–9, he claims that his discussion of ethical methods does not in fact depend on the idea that there are rational cognitions of ultimate ends (see also 77 n). All that is required is ultimacy, whatever the source of the end. Here we see a recognition in Sidgwick of the independence of at least some central issues in metaethics from normative or first-order ethics. Humeans, then, can speak of categorical imperatives, as long as they do not characterize them as rational.

Later, however, Sidgwick argues (against a Humean egoist) that, even if an ultimate end is not prescribed by reason, there is still a rational requirement to take the means to that end, which even a Humean must accept (36–8).²⁴ When your doctor says, ‘If you wish to be healthy, you ought to get up early’, she is not saying merely, ‘Getting up early is a necessary means to your being healthy.’ Her statement implies the *unreasonableness* of refusing to take the means to some acknowledged end. Sidgwick’s thought here seems to be that the Humean must accept that there is a categorical rational

22. See 35: ‘I at least do not know how to impart the notion of moral obligation to any one who is entirely devoid of it.’

23. This plausible approach to normativity has been developed at length by Parfit (2011: vol. 2, pt. 6).

24. See Nakano–Okuno 2011: 53–4. Sidgwick here anticipates Christine Korsgaard’s anti-Humean argument in her ‘Skepticism about Practical Reason’ (1986: 11–13). Deigh (1992: 252–3) sees the health example as problematic, in that there is no apparent conflict here between reason and passion to provide introspectible evidence of reason as motivational. But Sidgwick’s argument can be seen as independent of the earlier argument based on such conflict (23–4). And anyway the health example seems to provide an excellent illustration of the kind of case in which someone may be tempted not to take the means to their end while recognizing the irrationality of failing to do so.

requirement to take the means to one's end (see Prior 1963: 39).²⁵ And if there is this categorical requirement, then why should there not be categorical requirements to adopt certain ends in the first place?

Humeans, however, need not accept any rational requirement to take the means to one's ends (see Deigh 2004: 178). Just as there is nothing unreasonable in preferring the destruction of the world to the scratching of my finger, so there is nothing unreasonable about not taking the means to one's ends. Indeed, it can plausibly be argued that this was Hume's own view, since the means-end 'unreasonableness' he speaks of in the *Treatise* (2007: 2.3.3.6) is merely a mistaken judgement about causes and effects. As he puts it: 'a passion must be accompany'd with some false judgment, in order to its being unreasonable; and even then 'tis not the passion, properly speaking, which is unreasonable, but the judgement'.²⁶

3 Moral Motivation

Sidgwick sees ethical cognition and motivation as closely linked.²⁷ Near the beginning of the *Methods* (4–6), he raises the question why, if the aim of ethics is to tell us what ought to be and in particular what we ought to do, philosophers should concern themselves at all with psychology. After all, physicists get on quite well without discussing the faculty of sense perception. Sidgwick's first answer to this question appeals to the practical aim of ethics.²⁸ We cannot help believing what we see to be true, whereas we often end up doing what we know to be wrong on the basis of irrational desire. So it is important to understand the relation between moral judgement and the will.

25. It is not clear whether the Humean is meant to be committed to a narrow-scope requirement, so that if I have the desire to be healthy, there is a categorical requirement on me to get up early, or whether the requirement is wide-scope—so I am required, if I have the desire to be healthy, to get up early. The former is quite implausible, and leads to the 'boot-strapping' of reasons into existence (see Kolodny 2005: 512). So the more charitable interpretation would be in terms of a wide-scope requirement (see Broome 2007). This is probably how Sidgwick saw it, since otherwise it is not clear why a Humean must be committed to a categorical imperative to take the means to one's ends.
26. Note that Hume's phrasing would allow him to describe a passion or indeed a person as 'unreasonable', as long as he were not speaking 'properly', i.e. strictly.
27. For a persuasive argument that the 'fundamental distinction thesis' is primary in Sidgwick's account, and not merely part of an argument for a Kantian view of motivation (as argued in Shaver 2000), see Phillips 2011: 22–7.
28. Note that, though Sidgwick claims to be impartial in so far as he is not promoting any particular substantive ethical position, his aim is ultimately practical: to help us find out how we should act.

He then goes on to note the frequency with which people ask: ‘Why should I do what I see to be right?’ One might see the question as empty, since in the end it can be answered only by reference to some further ethical principle about which the same question can be asked. But we also need to explain why the question is so persistent. One explanation is that, since we are moved to act by desires *independent* of moral judgement, those asking this question are seeking an answer that in itself will prompt an overriding desire to do the action. Sidgwick allows that this explanation is sometimes true. But he also claims that, often, when people ask this question they assume that they are *already* determined to do whatever argument shows them to be reasonable, even if it goes against their non-rational inclinations. The main explanation for people’s asking the question so often is the variety of mutually inconsistent principles adduced in answer to it.

Sidgwick believes that the nature of this moral determination or ‘moral sentiment’ does not matter in ethics (76–7). Nevertheless, in a characteristically insightful and pithy discussion of moral phenomenology (39–40), he describes how the emotional aspect of the moral impulse varies according to the views and character of the person concerned. So an egoist may be moved by cool self-love, a utilitarian by sympathy. If one sees reason as external, one will experience a sense of reverence for authority, while if one identifies one’s self with reason, reverence will become self-respect. Different again are the emotions of admiration inspired by an aesthetic conception of virtue. What really matters, however, is that the sentiment in question, along with the apparent cognition of rightness common to all these experiences, operates successfully, even if it is in fact, say, a desire for the pleasure of acting rightly or to avoid the pains consequent on acting wrongly. This suggests that what makes a moral sentiment rational is not so much its content, but its origin as a response to a rational judgement.

Later (34–5), Sidgwick strengthens his claim about the link between moral judgement and motivation. When I cognize or judge ‘I ought to ϕ ’, he suggests, if I am using ‘ought’ in the narrower sense in which it implies ‘can’ and if I am rational, then this cognition will produce a motive to action—though it is not always overriding.²⁹ He goes on (see Kant 2002: 4.414):

29. Sidgwick occasionally uses language which might suggest that, on his view, motivation emerges from a standing desire or disposition to do what we recognize to be reasonable; but it is clear that he thinks that such cognitions produce motivation in their own right on particular occasions; see Shaver 2006: 6. Shaver also points out that, were Sidgwick to rely on standing desires, there would be no disagreement about motivation with Hume (as long as those desires themselves were not the result of rational cognition).

In fact, this possible conflict of motives seems to be connoted by the term ‘dictate’ or ‘imperative’, which describes the relation of Reason to mere inclinations or non-rational impulses by comparing it to the relation between the will of a superior and the wills of his subordinates. This conflict seems also to be implied in the terms ‘ought’, ‘duty’, ‘moral obligation’,³⁰ as used in ordinary moral discourse: and hence these terms cannot be applied to the actions of rational beings to whom we cannot attribute impulses conflicting with reason. We may, however, say of such beings that their actions are ‘reasonable’, or (in an absolute sense) ‘right’. (See 217)

Sidgwick appears to be assuming that a command can be made (a) only by a superior to an inferior, and (b) only when that inferior may be motivated to act in a way that conflicts with that command (see 41; 81). This is questionable (see White 1992: 317). Consider *modus ponens*. Many will say that, if *anyone* (regardless of their ‘status’ as superior or inferior) has sufficient reason to believe that *p*, and that if *p*, then *q*, then they are *required* by reason to believe *q* and ought to do so. Whether they have any disposition or even ability to believe otherwise seems irrelevant. The laws of logic, then, apply to all beings, so that Sidgwick’s distinction between ‘ought’ and what is “reasonable” or (in an absolute sense) “right” is unnecessary. Even if God is superior to all other beings, knows what is right, and cannot help but do what is right, we may, if we wish, say that he ought to act in this way or that he has a duty to do so.

But this raises the question whether we might not avoid the notion of ‘ought’ altogether, and prefer Sidgwick’s broader notion of what is reasonable (that is, of what one has overall reason to do).³¹ In general, Sidgwick retains more basic common-sense moral concepts than he needs, and as a result sacrifices clarity as well as having to spend a certain amount of time elucidating and defining them, and plotting their mutual relations.³² Sidgwick’s own understanding of ethics is stated clearly at 77, where he begins to summarize the results of the three preceding chapters:

The aim of Ethics is to systematise and free from error the apparent cognitions that most men have of the rightness or reasonableness of conduct,

30. At 218–19, Sidgwick discusses the suggestion that ‘duty’ and ‘obligation’ in particular imply a relation to some other person or persons, which may not be the case with ‘right conduct’ or ‘ought’. He correctly rejects it on the basis of common usage, instancing the duty of veracity.

31. Of course, ‘ought’, like ‘should’, can be interpreted in terms of overall reason. But its use may still cause confusion.

32. For the view that Sidgwick is sufficiently conceptually parsimonious, see Hurka 2003: 613–14. For the view that he should have done more plotting, see Bradley 1935: 72–3 n. 1.

whether the conduct be considered as right in itself, or as the means to some end commonly conceived as ultimately reasonable.

Even here the reference to ‘rightness’ can be eliminated, on the assumption that the ‘or’ is expegetic.³³ As Schneewind puts it (1977: 228), for Sidgwick, ‘[t]he point is to determine how we are to decide what it is ultimately reasonable to do, regardless of whether the ultimately reasonable is called “moral” or not’.³⁴ Parsimony requires any philosopher to ask her questions using as few concepts as possible, *ceteris paribus*, and indeed to introduce new concepts into her answers to those questions only when necessary. It may have turned out that Sidgwick could not state his own position without using the notion of ‘ought’. He might, for example, have ended up as a ‘dogmatic intuitionist’ with a list of moral duties. But the views he himself finds most plausible—egoism and utilitarianism—are best stated in terms of ultimate reasonableness, with no reference to ‘ought’ or morality at all (see Crisp 2006: ch. 1; Parfit 2011: 1.453 n. xl; forthcoming). It is, of course, true that Sidgwick’s method will require him to *mention* (rather than use) the word ‘ought’ in explaining normative intuitionism. But he could have pointed out the apparent flaws in that position without any serious enquiry into the logic of the word ‘ought’, and so avoided issues such as whether (the moral) ‘ought’ is to be understood in terms of punishment or whether ‘ought’ has a broad and a narrow sense.

An alternative strategy would have been to begin the *Methods* with a briefer discussion of common-sense morality, clarifying its essentially moral

33. This move would remove another common source of misunderstanding. Paul Hurley (2009: 11–12), for example, takes Sidgwick’s claim at 23 that ‘wrong conduct is essentially irrational’ to be evidence that he believed that ‘if something is morally prohibited, then you shouldn’t do it’, where the notion of a moral prohibition is the same as that found in common-sense morality. Hurley suggests (23) that ‘consequentialists reject a consequentialist theory of *rational* requirement, upon which an agent has decisive reasons to perform some action just in case its performance leads to the best state of affairs, and explicitly defend only a consequentialist theory of *moral* requirement’. On this view, Sidgwick’s utilitarianism would not count as a form of consequentialism.
34. Of course, Sidgwick had other aims as well: this is what I have been calling his ‘central’ aim. Earlier, Schneewind has claimed that Sidgwick’s basic notion is that ‘of a demand made by reason on action, or, more generally, the notion of a requirement which our own rationality presents to our desires and volitions’ (1977: 221). But there is no need to use the language of demands or requirements, or to introduce the notion of rationality, to elucidate the notion of an ultimate reason. For a good example of how Sidgwick might avoid Humean objections to his conception of moral judgements by focusing on judgements of reasonableness instead, see Shaver (2006: 11). He might also be able to sidestep the question of whether ethics should be seen as ‘imperative’ or ‘attractive’ (see 105–6; some of these issues are illuminatingly teased out in White 1992).

concepts only as required, and demonstrating its need for support by more fundamental principles. Then that supporting principle—utilitarianism—could have been discussed, like egoism, largely in terms of reasons alone. It is true, of course, that there would be a need to provide some account of reasonableness itself, and to plot the relations of that concept to certain others. But since reasonableness—in Sidgwick’s ‘overall reason’ sense—is more basic, and in certain ways simpler, than less fundamental concepts employed within certain domains of practical rationality, such as morality, these tasks would have been significantly more tractable than that of analysing the moral ‘ought’ and other moral concepts.³⁵

Let me now return to Sidgwick’s view of the relation between rational cognition and motivation. In one sense, Sidgwick might agree with Hume that reason is inert, in so far as it does not motivate even a rational agent on its own.³⁶ Rather, it gives rise to a sentiment which may then lead the agent to act. But in rational beings ‘as such’, recognition of a dictate of reason will produce such a motive. Sidgwick holds, then, some form of what has come to be known as *motivational internalism*, according to which there is a close link between ethical cognition and motivation. But it is, I suggest, a weak form of that view. In the summary at the beginning of 1.6, Sidgwick claims that apparent cognitions (‘apparent’ because they may be mistaken and hence not knowledge) are *normally* accompanied by moral sentiments. This suggests that he allows for cases in which a cognition—perhaps even a genuine one—does not give rise to any motivation. But Sidgwick will describe this as a case of irrationality, in so far as it is not an appropriate response by a rational being ‘as such’ to a rational demand (see Shaver 2000: 263; 267; 2006: 7).³⁷ This interpretation finds further support from Sidgwick’s claim at the

35. This approach would also enable Sidgwick to establish more clearly a direct link between reasons and well-being (see Crisp 2006: ch. 1). The reason I have not to hurt you, for example, is that I have a reason not to cause suffering (a reason that can be stated without using any special moral concept), not that I have a reason not to break some alleged moral directive or other. That is to say, Sidgwick might have offered a version of Williams’s ‘one-thought-too-many’ objection (1981b: 18), understood at the normative rather than motivational level, against all those theories that insist on grounding reasons in non-welfarist or moralistic notions.
36. For the view that Sidgwick rejected the Humean idea of reason as inert, see Deigh 1992: 242–4. Note, however, Sidgwick’s claim at 23 that ‘we do not conceive that it is by reason alone that men are influenced to act rightly’, an element of the ‘common view’ which he does not appear explicitly to reject (see Shaver 2006: 4–5). As Shaver points out, the first sentence of 25 implies that Sidgwick sees the difference between him and Hume to concern how reason influences desire.
37. ‘Rationality’, then, in these contexts is not to be understood procedurally or in terms of consistency ‘internal’ to an agent’s beliefs, desires, or actions. Sidgwick has a more substantive

end of 78 that, in a case in which I believe I ought to ϕ but lack any motive to do so, I cannot ordinarily see that lack of motive as a reason for not ϕ -ing.

4 Principles and Methods

Sidgwick suggests two ways in which ethics might be conceived (2–4; see 228). The first is the ‘jural’ view (see CD I.1), according to which ethics seeks the ‘true Moral laws or rational precepts of Conduct’, while the second—which we might call the *teleological* view—suggests that ethics is an inquiry into the ‘True Good’ and the method for achieving it. The jural view is said to be more prominent in modern ethical thought because such inquiry as there is into the ultimate good is directed at finding the actions that will lead to it. Sidgwick’s point here is hard to grasp, since he includes within his definition of the teleological view inquiry into the means of attaining the good. But he goes on to make a further claim, that the ‘intuitional view’, according to which conduct is right when in accordance with certain principles, cannot easily be understood as teleological.³⁸ Rather, it is concerned with explaining people’s moral duties independently of their own good or happiness.

As Sidgwick himself realizes, it seems that a proponent of the intuitional view could adopt a teleological position by claiming that a person’s ultimate good in fact consists in her performing her duty. But Sidgwick says that modern proponents of the intuitional view, and Christian thinkers in general, do not take such a view (see 391–2). This raises several important methodological questions. First, if a position reconciling the intuitional view and teleology was held as part of common sense in earlier times, why should it not be discussed now? Second, even if it has *never* been part of common sense, then why is that sufficient justification for our refusing to discuss it, if it has been taken seriously by past philosophers³⁹ or has some independent plausibility?

notion in mind, in which rationality consists in appropriate response to reasons. See Shaver (2006: 3), who elucidates Sidgwick’s conception of irrationality with the notion of going against the balance of reasons and (hence) being open to rational criticism. For modern variants of Sidgwick’s version of internalism, see e.g. Korsgaard (1986: 11); Smith (1994: 61).

38. As I understand Sidgwick here, he is speaking not only of ‘dogmatic’ intuitionism, but of any view that focuses initially on the rightness of actions rather than the ultimate good.

39. It was not an unusual view among ancient philosophers, including Aristotle, who claims that the human good consists entirely in virtuous activity, and that virtuous activity is constituted by doing one’s duty (1894: 1098a16–17; 1106b14–23; see Crisp 1994).

The subject of philosophical ethics, Sidgwick says, is the ‘methods’ of ethics, where a method is ‘any rational procedure by which we determine what individual human beings “ought”—or what it is “right” for them—to do, or to seek to realize by voluntary action’ (1–2; see 8; CD 2.3.1). Such a ‘procedure’ need not be a process. Sidgwick allows that a version of intuitionism, according to which we have immediate insight into the rightness of certain actions, is proposing a method (4). Nor is Sidgwick to be understood as suggesting that only actions matter in ethics, and not, say, the feelings or the characters of agents. Indeed, he elsewhere allows that the common-sense conception of virtue includes the emotions (222–3) and that ethics should construct ideals of character (393). But discussion of such topics is significant only in so far as it is related to the primary question of ethics—how we should act, that is, what we have overall reason to do. Like Aristotle (1894: 1103b26–9), Sidgwick sees philosophical ethics as essentially practical (1; 4).

The definition of method just provided leaves it open whether Sidgwick is including ethical theories, which advocate certain basic normative principles, as among the ‘rational procedures’ he has in mind. It appears not. Consider the view that God has implanted in us knowledge of certain apparently non-utilitarian common-sense rules, such as the rule that we should keep promises, because these rules are the best way to promote the utilitarian end of general happiness. According to Sidgwick, this view constitutes a rejection of the *method* of utilitarianism, though not of the utilitarian *principle* (85). Ethics, however, as Sidgwick points out, is ‘sometimes considered as an investigation of the true...rational precepts of Conduct’ (2–3), and he himself implies that we are interested in the principles that determine which conduct is ultimately reasonable (5–6; see 14). In other words, philosophical ethics is an inquiry into what grounds or justifies our actions and any decision-procedure we adopt, and so we might wonder why Sidgwick emphasizes methods rather than ultimate principles.⁴⁰ His book should perhaps have been titled *The Ultimate Principles of Ethics*, those principles each being a different statement of our ultimate reasons for action; and if it had been so titled, some of the lack of clarity introduced by his focusing on methods might have been avoided.

At 83, Sidgwick notes that, by adopting some plausible assumption or other, one can connect almost any method with almost any ultimate principle. But

40. Schneewind (1977: 194) plausibly suggests that Sidgwick was influenced by analogies between scientific and ethical methods.

he limits his own discussion to those methods ‘logically connected with the different ultimate reasons widely accepted’ (78). In other words, he is looking for certain commonly alleged ultimate reasons for action (such as the promotion of happiness), and the decision-making procedure that most naturally rests on each such reason. For example, egoistic hedonism consists in the acceptance of the agent’s own happiness as ultimate reason or end, and its method will be the attempt by the agent to maximize her own happiness. But since method and end can come apart—as in the case of Sidgwick’s own utilitarianism—one might wonder why Sidgwick appears to require that, before any ultimate reason can be given a place in his discussion, its related method must be present in common-sense morality.⁴¹ Imagine that common-sense morality made no room for utilitarian decision-making. The utilitarian principle could still capture what justifies the common-sense method or decision-procedure. Further, his emphasis on method over principle can lead to philosophical distortion. The focus on method may explain, for example, why Sidgwick is so ready to find utilitarianism within common-sense morality, because people sometimes decide what to do by trying to work out what would do the most good from the impartial point of view. This is evidence merely of *a* principle of beneficence, alongside others, and that is a far cry from utilitarianism as usually understood, according to which this is the only ethical principle.⁴²

41. Sidgwick believes that the principles lying behind the three methods he will discuss, and indeed those three methods themselves, *are* all found in common sense (14). But 421 makes it clear that what he means by this is that, for example, the principle of rational benevolence is *one* principle among others in common-sense morality, such that, when there is no conflict with other principles, it can be applied in the same way as utilitarianism proper, construed as a monistic position. 421 also provides some defence against Hurka’s charge against Sidgwick of ‘not clearly grasping the concept of a prima facie duty’ and hence not seriously considering ‘moderate’ deontology, the principles of which can be overridden (2014: 135; see Roberts 1969: 73). Ross’s notion is of course more carefully articulated, and—as Hurka shows—Sidgwick did not always have the distinction between absolute and moderate deontology clearly enough in mind (see CD, chapter 1 n. 43; CD 7.2.3), but he did understand the basic difference between an absolute principle and one which—not being ‘supreme’—is in potential conflict with others and hence overridable or open to limitation by them. His reason for rejecting moderate deontology is not that he fails to recognize the very possibility, but that it is imprecise, inconsistent, and so on. I do, however, agree with Hurka that Sidgwick’s failure to take moderate deontology seriously was a significant error on his part (see Broad 1930: 222–3).

42. See Calderwood 1896: 340; CD, chapter 1 n. 41. On one perhaps slightly uncharitable explanation of the focus on common-sense methods, Sidgwick knows his argument is heading in the direction of utilitarianism rather than dogmatic intuitionism, and recognizes that his task will be easier if all he has to do is persuade those attracted by dogmatic intuitionism to drop the non-utilitarian principles in common-sense morality and then seek a philosophical intuitionist foundation for what remains, which would then in effect be utilitarianism. See Hurka 2014: 149–50; Shaver 2014: 186–93; 200.

5 Sidgwick's Three Methods

According to Sidgwick, we continually inquire into what is ultimately reasonable because different and incompatible principles are present in common practical reasoning (6), and an answer given in terms of one of them will appear suspect from the point of view of the others. Which principles does Sidgwick have in mind?

The first set of principles, which will provide the source for normative (or 'dogmatic') intuitionism, comprises those that constitute the morality of common sense (7–8): the rules of prudence, justice, veracity, and so on. Sidgwick suggests that common sense ordinarily sees these rules as binding in any particular case, independently of the consequences of the action in question or its alternatives (see 96–7). By this, Sidgwick should not mean that, according to common sense, its rules apply *whatever the consequences*.⁴³ So common sense may allow that in unusual cases an injustice may be permissible or even required to secure some great good. But in ordinary cases it will hold that, say, the keeping of a promise is required because of the nature of the action in question and not because of the bad consequences of breaking the promise or of the good consequences of the practice of promise-keeping in general. There is a difficulty, however, in that the position Sidgwick will contrast with intuitionism—utilitarianism—can be understood in such a way that all the moral weight is placed on the nature of an action rather than its consequences. This is because actions are open to various descriptions. Consider the action of *maximizing happiness*. A utilitarian will see such an action as absolutely required in *every* case (not just ordinary cases), regardless of the consequences of performing that action. So I am required by utilitarianism to maximize happiness even if, for example, one consequence of my doing so is that an innocent person dies or is tortured.

Moving to the other two 'methods' in his triad, Sidgwick then notes (8) that many utilitarian thinkers see these common-sense rules as mere means to the general happiness of humanity or sentient beings, and contrasts this view with the principle of prudence which rests on the postulation of the

43. Though he sometimes does say things to suggest that this is what he means: see e.g. his reference to 'unconditional rules' at 78 (see 96), or the absolutist interpretations of the common-sense principles of veracity and justice at 20 (see Hurka 2014: 135–7). Nevertheless, Sidgwick often takes it for granted that the principles of common-sense morality may conflict.

individual's happiness as an end and provides the basis for egoism. He goes on to point out that people often adopt ultimate ends independent of happiness, such as fame, but suggests that this poses a potential problem for ethical theorizing if a different method has to be constructed for each end. It is not obvious, however, why this should be so, since the differences may be superficial. Compare the 'method' that involves the maximization of general happiness with that which requires the maximization of certain goods other than happiness. It is true that the actions resulting from the application of each method might be quite different; but the finding of the means to the different ends is a technical matter, and does not rest on any deep philosophical difference. Sidgwick anyway goes on to suggest that common sense will not, on reflection, validate such ends as rational. Rather, fame will be seen as valuable only in so far as it is a source of happiness to the famous individual, a sign of excellence or the achievement of some benefit, or a spur to further achievement.

By choosing fame as his example of a common end, Sidgwick has loaded the dice against those proposing goods other than happiness or excellence as worthy of pursuit—knowledge, accomplishment, or friendship, for example, construed as good in themselves and not in so far as they constitute human excellence (see Irwin 2009: 450; 459). Here (9) he merely asserts that, on the face of it, the only two goods with a strong and widely supported claim to being rational ultimate ends are happiness and excellence (i.e. the achievement, or partial achievement, of some ideal notion of human perfection). The reference to 'strength' here is important, and implies that there may be widely supported claims on behalf of other goods which Sidgwick sees as weak.⁴⁴ This is an early example in the *Methods* of a potential clash between common sense and Sidgwick's own intuitions.⁴⁵

Sidgwick does provide an argument for restricting his discussion of ultimate reasons to those based on the goods of happiness or excellence, and those based on duty as prescribed by unconditional rules (78). He suggests

44. As noted by Schultz (2004: 165), in earlier editions Sidgwick was readier to admit that he may have devoted insufficient attention to perfectionism.

45. See also 10 n. 5, where Sidgwick considers the suggestion that perfection as an end might be extended to inanimate objects, but suggests that 'reflection...shows' that qualities such as beauty cannot be good in themselves, independently of the perfection or happiness of sentient beings. He refers forward to 1.9, where we read, for example: 'no one would consider it rational to aim at the production of beauty in external nature, apart from any possible contemplation of it by human beings' (114). It is this passage which prompted Moore's famous argument in favour of the view that a beautiful universe has value in itself (1903: 83–4).

that this tripartite distinction mirrors what appear to be the most fundamental distinctions we use in understanding human life: that between the conscious being and the stream of experience, and that within the stream of experience between action and feeling. Perfection is put forward as the ideal goal for a human being, considered diachronically, while duty concerns what should be done, and happiness is a matter of feeling. But on the face of it, even if we take some version of this trichotomy to be fundamental, it is hard to see why it must map onto the distinction between excellence, happiness, and duty in the way Sidgwick suggests. Happiness or duty could be said to be the ideal goal; perfection, as moral excellence, concerns both action and feeling. Why, anyway, should we see this distinction as fundamental? Why not that between, say, mind and body? And other goods could be said equally plausibly to map onto the distinction. So beauty, for example, could be said to be a quality related primarily to the conscious being, as opposed to her actions or feelings.

Sidgwick allows for happiness to provide the foundation for two different methods, egoism and utilitarianism, depending on whether it is the agent's own happiness or general happiness which is to be sought. We might expect the same distinction between self-regarding and impartial perfectionist principles, but Sidgwick claims that past philosophers have always assumed, as far as 'moral excellence' is concerned, that promoting the virtue of others is never in tension with the agent's own virtue.⁴⁶ First, however, even if Sidgwick's historical claim is correct, the question is whether common sense might accept the possibility of such a tension. It certainly seems that it should. All we have to imagine is some case in which, by refusing to act wrongly, I bring it about that two or more others act wrongly to the same degree.⁴⁷ Second, his focus on moral excellence in particular enables him to ignore examples of sacrifice of the agent's own non-moral excellence for that of others—such as, perhaps, the return of the philosophers in Plato's *Republic* into the cave (2003: 519c8–520e3). Sidgwick does provide an

46. See also 404–5, where Sidgwick attributes the common ancient view that it can be 'good for' an agent to sacrifice her happiness for others to a confusion between what it is reasonable for an agent to desire when she takes only her own existence into account and what it is reasonable to desire from the point of view of a larger whole, or to a profound faith that such a conflict could not exist. He explains the popularity of the hedonist conception of well-being as the result of the drawing by Butler and others of a clear distinction between self-love and conscience.

47. Of course it might be claimed that this is impossible, since it must be virtuous to prevent others' acting wrongly. But this is a substantive not a conceptual claim, and it could be plausibly denied by those who believe that duties are 'agent-relative'.

argument for the restriction of excellence to moral excellence, which is that virtue is ‘commonly conceived’ as not only the most important component of excellence, but a component lexically prior in value to any other component. This, however, leaves open the possibility of a different conception, perhaps equally ‘common’. Consider a case in which I am in a position greatly to advance knowledge, accomplishment, and friendship, either in my own life or generally, but only through committing some minor moral misdemeanour such as telling some not terribly important lie. I suspect that many would permit the lie and perhaps even see it as rationally required.

Sidgwick’s restrictive conception of excellence leads him to see perfectionism as merely a special type of intuitionism (the view discussed above, according to which common-sense moral principles are ultimately reasonable). But here we see another philosophical distortion arising out of his focus on method rather than principle. Even if the views are extensionally equivalent, they should not be identified.⁴⁸ According to intuitionism, the reason I should keep my promise, for example, is that promises should be kept. According to perfectionism, however, the ultimate reason here consists in the promotion of perfection.

The upshot of my discussion so far, then, is that Sidgwick should have considered, in addition to the three methods he does discuss, a further two: egoistic and general perfectionism (see Irwin 2009: 449).⁴⁹ Further, it is tempting to think that Sidgwick may be allowing his own intuitions to ‘filter’ the deliverances of common sense at an early stage in the argument.⁵⁰ A footnote to the sentence in which he says he will treat perfectionism as a

48. See Schneewind 1977: 202–4; Irwin 2009: 450. Schneewind suggests at 204 that the link Sidgwick drew between perfectionism and (dogmatic) intuitionism led to a failure to find any non-perfectionist principle with which to connect the intuitionist method. But it is clear that Sidgwick saw duty as providing an ultimate reason, and hence as grounding an ultimate principle, which is non-teleological and hence independent of both happiness and excellence. See e.g. 78; also 7–8. Sidgwick’s error, then, is to characterize perfectionism in terms of intuitionism, not vice versa.

49. Indeed, ‘happiness’ theories might be said themselves to split into different categories, depending on whether their conception of happiness is hedonistic, preference-based, or a form of ‘objective list’ theory in which the good is constituted at least in part by certain non-hedonic goods which are valuable independently of their fulfilling a desire (see Parfit 1984: App. I). This latter view is, to speak strictly, not a version of perfectionism, since it makes no reference to perfection or human nature in its account of the good. It might also be thought that Sidgwick should have considered positions between egoism and impartial utilitarianism, according to which I should give priority to, say, my own children. Here I think Sidgwick could plausibly claim that his conception of dogmatic intuitionism is broad enough to cover such a position.

50. For a further example, consider 78–80, where Sidgwick ignores the possibility of a divine command view which does not collapse into a happiness- or excellence-based position.

form of intuitionism refers us to the argument in 3.14 against non-hedonistic conceptions of the good. Again, then, Sidgwick's methodology is perhaps not quite as he suggests.

The most plausible explanation for Sidgwick's discussing egoism, common-sense morality or intuitionism, and utilitarianism is that these are the three views *he* finds most plausible:

When I am asked, 'Do you not consider it ultimately reasonable to seek pleasure and avoid pain for yourself?' 'Have you not a moral sense?' 'Do you not intuitively pronounce some actions to be right and others wrong?' 'Do you not acknowledge the general happiness to be a paramount end?' I answer 'yes' to all these questions. My difficulty begins when I have to choose between the different principles or inferences drawn from them. (14)

Further evidence for this suggestion may be found in what Sidgwick says at 10 about the possibility of ends intermediate between the happiness of the agent and general happiness, such as the happiness of one's family, nation, or race: 'any such limitation seems arbitrary, and probably few would maintain it to be reasonable *per se*, except as the most practicable way of aiming at the general happiness, or of indirectly securing one's own'. This claim about the views of others is highly implausible, and is an especially clear case of Sidgwick's assuming too quickly that others will, on reflection, agree with him. Given a choice between utilitarianism and a principle that allows or even requires the agent to give some priority to her own children, it is not unlikely that the majority will prefer the latter.

As we saw above, Sidgwick's views on the autonomy of ethics also led him to restrict the scope of philosophical ethics as he conceived it. In the end, however, the doubts I have raised about Sidgwick's choice of three methods do not, I believe, undermine his project, broadly construed (see Nakano-Okuno 2011: 42). Moral theories can plausibly be distinguished as broadly consequentialist on the one hand, and non-consequentialist or deontological on the other.⁵¹ And the main opposition to moralism is, of course, philosophical egoism, which, though it is not currently much discussed in contemporary ethics, remains one of the most powerful and attractive normative theories. Indeed, Sidgwick could plausibly have started with these three methods, and gone on to explain how different accounts of the good will give rise to different forms of egoism and utilitarianism in

51. In Crisp 2015a, I explain why virtue ethics is not a separate category of theory, but a form of non-consequentialist deontology. See White 1992: 313.

particular. The arguments for his basic triad are problematic; but it stands independently of them.

6 Ethics and Politics

As in Aristotle's *Nicomachean Ethics*, the second chapter of Sidgwick's *Methods* addresses the question of the relation of ethics to politics. In the very first paragraph of the *Methods*, Sidgwick says that politics aims to explain which constitution is best and which 'public conduct' is correct in a governed society. So, though like ethics it is concerned with what ought to be rather than what merely is, its subject is not the individual but rather society. As Sidgwick makes clear at 15, this includes, in addition to the government itself, also the governed, in so far as they are in a position to exercise control over the government. Sidgwick's definition, as he realizes (18), removes anarchist options from the scope of politics. He would almost certainly have thought anarchism unworthy of consideration,⁵² but he does allow that 'politics' might be understood more broadly to cover all social relations (18 n. 2).

The mapping of the ethics/politics contrast onto that between the individual and the group might be criticized on grounds of methodological individualism, since any actions of a government or those influencing government or the constitution will be actions of individuals (15). So politics must be a branch of ethics. Sidgwick responds with an attempted *reductio*: on this view, ethics would have to cover all of the arts, including, say, navigation or medicine.

Sidgwick seems to be taking the objector to be committed to the view that all actions of individuals must be the concern of ethics, including, say, those of steering a ship or curing someone. But these actions are means to ultimate ends, and those ends are what lie within the scope of ethics. And, like Sidgwick himself, we should not want to understand politics purely instrumentally or technically. Imagine that the objector is a utilitarian. Her point is that, just as my ultimate reasons in the private sphere are determined by the greatest happiness, so the same will be true if I am acting as a member of a government or the governed. What Sidgwick might have suggested is that ethics can be extended into the political domain, where it may

52. At EP 169, he calls it a 'wild imagining', which does 'not concern us'.

concern itself in part with what should be done by actors other than individuals, such as governments or other groups.

But it will still be tempting to explicate claims about the obligations of such non-individual actors in terms of the obligations of individuals, perhaps restricted to certain times. So if I claim that the present government has a duty to alleviate child poverty, what I mean is that, in effect, these individual cabinet members now have a duty to do so, and, if there is a reshuffle, the new ministers will have that duty. This is not to rule out a distinction between ethics and politics, however. Politics could still be concerned with individual obligations in the public domain, or the obligations of political groupings, even if the obligations of groups are also to be seen as, ultimately, ethical obligations of individuals. Politics, that is to say, is a species of ethics.

Sidgwick goes on to note the close relation between politics, law, and morality (16–18).⁵³ Legal regulation should be in line with morality, so that no law requires anyone to act morally wrongly. Further, it will in many cases *determine* an individual's moral obligations. Consider the duty of justice to 'give everyone their own'. Exactly what is required of any individual by this will depend on the law of her state.

By 'law' here, we must mean 'law as it is' or *positive law*, Sidgwick suggests. Political theory, however, is concerned with *ideal law*, or law as it ought to be. Sometimes, the gap between ideal and positive law is so great that it may affect what I take to be my moral duty under the law as it is, but the extent to which this is so is 'vague and uncertain'. Consider the case of a slave-owner who has decided that the positive law allowing property rights in human beings should be abolished. She need not immediately free her slaves, perhaps even on the ground that this would not be best for them. Rather, she may treat them better, in the hope of setting an example to other slave-owners, and begin a campaign to improve the law regarding the treatment of slaves. In general, Sidgwick claims, the degree to which a political ideal affects moral duty will depend on the chance of realizing it and its urgency or the benefit of realizing it. These in turn will depend on which political method is selected, so that determining these issues themselves depends on politics rather than ethics. The nature of ideal law is, Sidgwick

53. See also the passage appended to the end of Sidgwick's discussion of utilitarianism and the morality of common sense (457–8). The passage was originally in 2.1, but Sidgwick had removed it from there in the sixth edition with the intention of incorporating it into book 4, the revision of which was interrupted by his death.

suggests, less directly within the purview of ethics. I take it that what he means is that the concern of ethics is what I should do, here and now, in the light of the law as it is. It may be that this will itself depend on the distance between positive and ideal law, in which case ethics will take ideal law into account. But otherwise it will not, and the determination of ideal law itself is a matter for political theory.

It is not clear why Sidgwick restricts the role of political theory to the working out of ideals only. For the sphere of 'what ought to be' covers not only ideals, but which non-ideal goals might be worth adopting in the light of probability judgements, feasibility constraints, and so on. Again, it is tempting to see political theory as merely that aspect of philosophical ethics concerned with politics. Consider Sidgwick's own utilitarian position. This view is quite clear on which laws are ideal: those which we find in the best history of the world, assessed from the utilitarian perspective. And, as in the case of Sidgwick's slave-owner, utilitarianism can also explain whether, if positive law deviates from the ideal, I should act as the ideal law would require, or in some other way.

In the second section of 1.2, Sidgwick discusses those Kantian theories that see ethical theory as concerned not with what we ought to do here and now, but with what ought to be the rules in a society of ideal human beings.⁵⁴ On this view, the relation between ethical theory and everyday practical questions is as indirect as that between political theory, as construed by Sidgwick, and such questions. Sidgwick rejects the suggestion that ethics should be modelled on a conception of geometry as concerned only with ideally straight lines, circles, and so on, noting first that geometers do offer accounts of irregularity as found in the world, and then that the primary ethical question of interest to most people is what they are required to do as things are.

The working out of an ethical ideal can, then, be only a preliminary to an account of our duty here and now. Sidgwick suggests that how useful it will be to develop such an ideal will depend on one's ethical theory, claiming that for intuitionists—who generally believe in universal and absolute principles—duty may be thought to be the same in both ideal and actual worlds (one is always required to tell the truth, say, whatever the state of one's society). Whether the intuitionists of Sidgwick's day were, or needed to be, absolutists in this sense does not at present require further discussion.

54. He has his contemporary Herbert Spencer's *Social Statics* especially in mind; see 18–19 n. 2.

For Sidgwick goes on to admit that even an 'extreme' intuitionist will allow that what exactly is required by various ethical principles will vary according to the state of current institutions.

What about utilitarianism (20–2)? Sidgwick does not deny that ideal theorizing may be helpful. But he points out how difficult it will be, since the boundaries of the possible are so unclear. Plato, for example, appears in the *Republic* to have seen war as inevitable; while in contemporary times some see permanence of affection between spouses as natural and so as necessary for the ideal, while others see it as contingent and prefer an ideal of 'free love'. In other words, though we may characterize the ideal as that in which the balance of pleasure over pain is maximized, it is very unclear which institutions and rules will be partly constitutive of that ideal. Further, even if we could, on the basis of fuller and more accurate historical knowledge of human development than we now have, predict which rules will in future maximize utility, it remains unclear what the implications of that ideal are for us here and now.⁵⁵

55. At 22 n, Sidgwick refers the reader to 467–71. For discussion of that passage, see CD 7.2.4.

2

Free Will

1 The Aim and Structure of 1.5

Sidgwick's chapter on free will follows a chapter on pleasure and desire, and precedes a chapter on methodology, which itself takes up themes from the third chapter of book 1, 'Ethical Judgments'.¹ The main aim of the chapter is to show that the issue of free will is of little importance for ethics, and Sidgwick explicitly says he will not himself offer any substantive view on it (66).² Since Sidgwick is keen to avoid the great philosophical difficulties in the debate, its ethical insignificance is, from the methodological point of view, welcome to him (57).

Interpretation of the chapter is made harder through its internal organization, which is as follows (references are to section and paragraph):

Section 1: Clarification: We should not identify free actions with either disinterested or rational actions.

The main issue of the free will debate: Could a wrongdoer have chosen rightly, even if the causal antecedents of her action had been identical? (See 2.6 of this outline below.)

Section 2: 1–5: The scope of the debate: Voluntary actions.

6: The main issue of the free will debate.

1. Sidgwick placed it where it is as part of his general discussion of the nature of ethics, though it might be that the overall structure of book 1 would have been more straightforward had he ordered the existing chapters as follows: 1, 3, 6, 2, 5, 4. This would have provided a trilogy of chapters on ethics, followed by a chapter on each of politics and free will, before a chapter on pleasure leading into 1.7 on egoism. Much of 1.9 ('Good') could have been included in that chapter, before 1.8 on intuitionism. Book 1 is also lacking an introductory chapter on utilitarianism. (Indeed, I myself think it would have been wiser for Sidgwick to omit 1.7–9, and to include any relevant material on egoism, intuitionism, and the good in later books.)
2. See his response to critics of the *Methods* at SFEC 36: '[N]o other aim but . . . removing misunderstandings could have induced me to recur to the ancient problem of the Freedom of the Will. I have no pretension of providing a theoretical solution of this problem; and, indeed, the first misunderstanding which I wish to remove is one which attributes to me such a pretension.'

7–10: Arguments for determinism.

Section 3: 1: The main argument for libertarianism. The irrelevance of the debate to ethics.

Libertarians exaggerate the importance of the free will debate.

2: The libertarian position on the significance of the debate.

3: Freedom is often understood in a compatibilist sense. Even when it isn't, neither libertarian nor (hard) determinist views have much significance for practical deliberation.

4–5: They lack significance for which ends it is rational to pursue, except in so far as the question arises of the reward for virtue in the afterlife.

6–7: The debate is largely irrelevant to predictions of the future.

8: Summary.

Section 4: 1–2: Determinism does have implications for notions of remorse, desert, and so on, but satisfactory utilitarian accounts of these notions, consistent with determinism, are available.

Section 5: 1–5: The scope of the will.

So this chapter contains important discussions of Sidgwick's conceptions of freedom, intention, the will, voluntariness, responsibility, desert, and other central ethical notions. Further, Sidgwick's own views on individual issues in the free will debate itself do of course emerge at various points. One of the most revealing aspects of the chapter as a whole is its illustration of how even in book 1 Sidgwick's own utilitarian or consequentialist view of normative ethics is already waiting in the wings, ready to be produced as required.³

2 Disinterested Action, Rational Action, and Kant

In the first section of 1.5, Sidgwick begins by defending his decision not to address the question of free will in discussions in previous chapters of rational and of disinterested action. Disinterested action is covered in the

3. As far as the structure of the chapter goes, perhaps the elements that cause the most confusion are Sidgwick's tacking on of the arguments for determinism to the end of section 2, and his placing of the discussions of remorse, and so on, into a separate section 4. It is again hard to resist the thought that a more straightforward ordering of the discussion was available. Sidgwick might have hived off 2.6–10, the first part of 3.1 on libertarianism, and 3.2 on the libertarian view of the significance of the debate into a self-standing section. Then might have followed a section explaining the irrelevance of the debate for ethics, including the rest of 3.1, 3.3–7, and section 4. I offer this suggestion not as a rather pointless objection to Sidgwick's chosen structure, but in the hope that it may enable the reader better to follow his argument in the chapter as it stands.

preceding chapter, and consists in an individual's acting without considering the effect of the action in question on the overall balance of pleasure and pain for herself. Sidgwick claims that, because such action is often instinctive, disinterested action cannot be identified with free action. As an example of such disinterested action, he offers the eating of food prompted by hunger (45). A little later in the chapter (61), Sidgwick defines an instinctive action as one caused by feelings of pleasure and pain, and not preceded by any representation 'in idea' of the bodily movements in question or their effects. For Sidgwick, intention consists in such representation (60), so by instinctive eating, he must mean a case in which an extremely hungry agent 'finds herself' eating, not having taken any conscious decision to begin doing so in the light of the kind of representation he here describes. As we shall see, this conception of intention is questionable, as is the restriction of the scope of judgements of responsibility to such conscious volitions. Such eating might, for example, be a case of culpable negligence by someone put in charge of the food laid out for some banquet, and so within the domain of the debate on free will and responsibility.⁴ Nevertheless, there seems no reason to deny Sidgwick's claim that such disinterested action is not to be identified with free action, if only because much 'interested' action appears to be free.

The equation of rational and free action, however, is worthy of greater attention, not only because of its greater intrinsic plausibility, but also because of its Kantian pedigree. Sidgwick's first point here appears to be an expression of compatibilism about rationality and determinism:

[R]ational action, as I conceive it, remains rational, however completely the rationality of any individual's conduct may be determined by causes antecedent or external to his own volition: so that the conception of acting rationally, as explained in the last chapter but one, is not bound up with the notion of acting 'freely', as maintained by Libertarians generally against Determinists. (57)

Sidgwick sees a rational action as one that consists in an appropriate response to reasons. If the agent's willing of such an action is itself entirely the product of causes antecedent to that willing, it remains the appropriate response,

4. Influenced by Aristotle's discussion of voluntariness (1894: 1114a3–10), Sidgwick might suggest that this individual wilfully chose to allow herself to become a glutton. But she might have been equally negligent earlier in her life, and never stopped to think about the direction of development of her character.

and so rational. Many are likely to see Sidgwick's conception of rationality as excessively objective, insisting, for example, that to be rational an agent must meet certain epistemic or procedural coherence requirements. It might even be objected that a completely determined 'action' is better described as a mere event, in which the exercise of the will is seen merely as another link in the causal chain leading to that event.⁵

Is 1.5 another result of Sidgwick's distinction between reasonableness or rightness (i.e., perhaps, rationality) and what is required by duty? For it might be thought that were he willing to extend his compatibilism to judgements about what we *ought* to do, he could make the case for the irrelevance of the free will debate to ethics without further argument. Indeed, his summary of the results of 1.5 at the beginning of the following chapter might even suggest that this is how he himself understood the strategy of 1.5:

Nor, again, is it generally important to determine whether we are always, metaphysically speaking, 'free' to do what we clearly see to be right. What I 'ought' to do, in the strictest use of the word 'ought', is always 'in my power', in the sense that there is no obstacle to my doing it except absence of adequate motive; and it is ordinarily impossible for me, in deliberation, to regard such absence of motive as a reason for not doing what I otherwise judge to be reasonable. (78)

But this would be to read Sidgwick as having made a move in the very debate he is scrupulously trying to avoid.⁶ This passage should be read in the light of the crucial paragraph (67–8) in which Sidgwick claims that the (hard) determinist will accept the common view that the 'can' in 'ought implies can' means 'can do if I choose'. Of course, this determinist will then go on to ask the question: 'But *can* I choose to do what I judge to be right?' Sidgwick's answer to that question is that he himself thinks he can. But he is not here using a libertarian intuition in an attempt to dislodge determinism. Rather, he suggests, determinism would make a major difference to ethical deliberation only if it often turned out that one could be sure that one would *not* choose what is right:

But I think the cases are rare in which it is even on Determinist principles legitimate to conclude it to be certain—and not merely highly probable—that I shall deliberately choose to do what I judge to be unwise.

5. 66 suggests that Sidgwick would have been prepared at least to entertain this objection.

6. At M1 51, Sidgwick confesses that he thinks sometimes as a libertarian, and sometimes as a determinist.

In other words, Sidgwick's position is that what really matters in practical deliberation is whether, if I choose to ϕ , I am indeed in a position to ϕ . Determinists may be correct that my choices are themselves determined; but if there is a chance that I may choose to ϕ , and a chance that if I so choose then I will ϕ , my deliberation need not be impeded by a belief in determinism. This is in a sense a form of compatibilism; but it is not one based on the Humean conception of freedom as lack of impediment to action. (More on 1.5.3 in CD 2.5.)

In expressing his compatibilism about rationality, Sidgwick implies that libertarians wish to defend a contra-causal account of freedom. He goes on to make clear that *some* libertarians—'disciples of Kant'—will in fact wish to identify freedom with rational action. But, he suggests, these Kantians are confusing two senses of freedom, which have to be carefully distinguished. In one sense, a free action is a voluntary action in which the agent follows the guidance of her reason and successfully resists the influence of 'appetite or passion'. Sidgwick acknowledges the phenomenon (indeed, it is what he mentions as the strongest evidence for libertarianism at 65), and is prepared to allow this usage of 'free', recognizing its protreptic potential (though also noting somewhat acerbically, and presumably with a nod to Hume's view of reason as slave to the passions, that, if one identifies oneself with one's passion or appetite, obedience to reason appears 'a servitude, if not a slavery' (58 n. 2)). But it is important to distinguish this sense of freedom from that central to libertarianism, which allows for the freedom to choose irrationally as well as rationally, as in cases of weakness of will in which a person prefers, say, indulging in sensual pleasure to preservation of her health.⁷ And this brings us to the central question in the debate between libertarians and determinists:

Can we say, then, of the wilful wrongdoer that his wrong choice was 'free', in the sense that he might have chosen rightly, not merely if the antecedents of his volition, external and internal, had been different, but supposing these antecedents unchanged?

7. It is worth noting a characteristically ingenious footnote on the subject of weakness of will (59 n). Sidgwick claims that we find it easier than the ancient philosophers to allow for full-blooded weakness of will because we draw a clear distinction between moral and prudential (self-interested) judgements. Since we then have the conceptual apparatus to describe action motivated by one of these concerns in opposition to the other, we can without difficulty conceive of action opposed to both. In my view, Socrates, Plato, Aristotle, and other ancient philosophers had a clear grasp of the distinction between morality and prudence; but they were certainly interested in closing the gap between the two, and this may well have had an effect along the lines of that suggested by Sidgwick.

Because of the confusion of these two senses of freedom, and because of the Kantian conception of causation as ‘not subject to time-conditions’, which Sidgwick finds both implausible and anyway unsuitable for discussion in a treatise on ethics, Sidgwick does not discuss the Kantian conception of free will in detail in his main text (58 n. 1). But he does append to the *Methods* an appendix on the subject, a partial reprint of an 1888 article in *Mind*. That appendix accuses Kant of the same confusion as his disciples between *good* or *rational* freedom on the one hand, and *neutral* or *moral* freedom on the other (see also GSM lec. 2 on Green). (Sidgwick notes that the first name in each disjunction brings out better the difference between the two positions.)

Sidgwick distinguishes both good and neutral freedom from what he calls *capricious* freedom—the Reidian conception of freedom as the capacity of acting without any motive at all—and claims that excessive focus on Kant’s rejection of capricious freedom has led readers to overlook the good/neutral distinction. Sidgwick provides strong textual evidence in Kant of both conceptions of freedom. The thief is free when he steals, yet freedom consists in the resistance of inclination and obedience to the moral law. The question, of course, is whether these two notions of freedom are somehow incompatible, and Sidgwick admits that a Kantian may reconcile neutral freedom with universal physical causation by bringing in the idea of a noumenal self not itself subject to such causation: the thief’s action is determined at the phenomenal level, but free at the noumenal. But, he continues, the Kantian will then be required to give up the idea of the ‘heteronomy’ of the will, and this is a serious loss,

since nothing in Kant’s ethical writing is more fascinating than the idea—which he expresses repeatedly in various forms—that a man realises the aim of his true self when he obeys the moral law, whereas, when he wrongly allows his action to be determined by empirical or sensible stimuli, he becomes subject to physical causation, to laws of a brute outer world. (App. 516)

It is unclear, however, why the Kantian should not give an account of heteronomy using exactly the good/neutral distinction which Sidgwick has accepted she may retain (see Paton 1971: 214–15; Skorupski 2000: 322). Realizing the aim of one’s true self is *good* freedom and autonomous. Failing to do so is to surrender an opportunity to exercise good freedom, and since

one is led by one's inclinations, it is a case of heteronomy. But one is still *neutrally* free, in so far as one *could* have chosen to follow the moral law. The real difficulties with the Kantian conception of freedom are in the transcendental idealism with which Sidgwick chooses not to engage. Once that is granted, both good and neutral conceptions of freedom can be imported into the account.

3 The Will

3.1 Voluntariness, Volition, Intention, and Responsibility

Sidgwick begins the second section of 1.5 with the claim that all methods of ethics restrict the predicates 'right' and 'what ought to be done', in their strictest senses, to voluntary action (see 426–7). There are two initial problems with this claim, one more serious than the other. The less serious problem is that some philosophers will be prepared to attribute rightness or wrongness to involuntary actions. Certain objective utilitarians, for example, will claim that some action is right because it led to the best consequences overall, even if it was entirely involuntary. (Whether Sidgwick's own view is in this category is a matter for later discussion. And whether the view is plausible is not the issue here.) The more serious problem is that nearly everyone, including most philosophers, will allow that agents can act wrongly through negligent *failure* to act voluntarily: more on this below. But we must surely agree with Sidgwick that voluntary action is standardly seen as a paradigmatic object of moral appraisal.

How, then, are we to understand voluntariness? Sidgwick suggests first that voluntary actions must be conscious, in the sense that the agent must be aware of them during their performance. He appears here to equate an action's being unconscious with its being 'mechanical'. But we can be conscious of certain mechanical actions, such as a knee-jerk reflex. Sidgwick later (61) extends his account to cover conscious actions performed without intention (that is, without the agent's representing to herself the nature of the action or its consequences), and gives as an example the class of 'instinctive' actions caused by sudden feelings of pain or pleasure. Instinctive actions are structurally analogous to mechanical ones, and so we may presume that Sidgwick intends that conscious mechanical actions also be excluded from the sphere of the voluntary.

Sidgwick then goes on to suggest that agents are never seen as culpable for entirely unforeseen effects of their voluntary actions. He mentions the possibility of negligence, but claims that ‘reflective persons’ will hold that such an agent is being blamed only ‘indirectly’, in so far as the negligence is the result of some ‘wilful’ (i.e. voluntary) failure of duty. This view of the reflective common-sense conception of negligence cannot be put down directly to Sidgwick’s own consequentialist leanings. Indeed, the scope for blame in consequentialism is limitless.⁸ Anything can be blamed, if blaming produces the best result. Further, even if we restrict attributions of blame to wrongness as ordinarily understood, an act utilitarian may coherently blame some agent for negligently failing to produce the best consequences (which may of course also increase the likelihood of her not similarly failing in future).

It may be, however, that Sidgwick’s consequentialism does lie behind his next claim, that:

the proper immediate objects of moral approval or disapproval would seem to be always the results of a man’s volitions so far as they were intended—*i.e.* represented in thought as certain or probable consequences of his volitions:—or, more strictly, the volitions themselves in which such results were so intended, since we do not consider that a man is relieved from moral blame because his wrong intention remains unrealised through external causes. (60; see 201–3)

There are several things to be said about this brief outline of Sidgwick’s philosophy of action.

- (1) We should not press the distinction between an action and its consequences against Sidgwick here. We should not object, for example, that we might wish to hold someone to account for murder, even though all she had willed was the *action* of murder itself, and not any *consequences* of it. For Sidgwick wishes to see any voluntary action as a ‘consequence’ of a volition—that is, a willing, or a decision to act.
- (2) Further, we should probably not see Sidgwick’s account as over-intellectualized, as if the only objects of moral appraisal were carefully considered decisions to act based on a careful analysis of the result of that decision’s being carried through. All that is required is a degree of

8. At 72 n. 2, Sidgwick claims that ‘we find it necessary to punish negligence, when its effects were very grave, when we cannot trace it to a wilful disregard of duty’.

awareness on the part of the agent of what she is doing or about to do. If I steal something on impulse, for example, that will count for Sidgwick as a voluntary action which I have decided to perform (61). But Sidgwick is right to claim here that common-sense morality tends to see deliberate wrongdoing as worse, other things being equal, than impulsive. In the case of impulsive actions, the ‘consciousness of personal choice of the result is evanescent’. One might offer a Humean explanation of this difference in approach (see Hume 2007: 3.3.1.4–5). When we praise or blame an action, we are really praising or blaming the person who performed it for having certain qualities that disposed them to perform it. An impulsive action may be less indicative of a person’s real character than an action carefully deliberated in the light of the agent’s values. Nevertheless, someone who wished to extend more broadly than Sidgwick the scope of common-sense moral appraisal might wish to claim that, even if consciousness of personal choice were entirely absent, such that the person acted unthinkingly and ‘found herself’ performing the action, blame might be as appropriate as in the case of an action emerging from deliberation. (See (5) below.)

- (3) But we should note the strongly cognitive nature of his conception of intention as mere representation, which appears to leave no room for intentions to play any guiding role in action.⁹ For Sidgwick, it seems, moral judgements concern—and concern only—the results which an agent believes will arise from her decisions.
- (4) We have already seen that Sidgwick’s account fails properly to explain responsibility for negligence. But his restriction to foreseen consequences also causes difficulties with voluntary action itself, once we recognize that the context of decisions is opaque rather than transparent. So when Oedipus made the decision to strike the man at the crossroads, we cannot plausibly say that he decided to strike his father. The striking of his father, which was to result from his decision, was quite unforeseen. And yet some will blame him for striking his father.¹⁰

9. For a more recent statement of a cognitivist view, see Velleman 1989: chs 2–3; for discussion and criticism, see Paul 2009: 3.

10. Of course, others will see such blame as unreasonable, especially if it is not directed at some alleged negligence on the part of Oedipus. But it does occur, as a third-person analogue of what Bernard Williams (1981a: 27–8) calls ‘agent-regret’ in the first-person case. (Perhaps it would be better to speak here of ‘moral resentment’ rather than blame, though the two are bound to shade into one another.)

- (5) Sidgwick's account leaves out also many other objects of moral appraisal which are independent of decisions and acts. Agents are often blamed for their feelings (such as excessive or unjustifiable anger), their desires, wishes, or hopes (e.g. that someone suffer terribly), or even their beliefs (as in the case of racism).
- (6) Finally, and relatedly, Sidgwick's denial of so-called 'moral luck' at the end of the passage now seems highly questionable. It is true that one underlying principle in common-sense morality is that blame should be distributed 'fairly', in the sense that no agent should be blamed more or less than any other merely because of 'bad' or 'good' moral luck. If we both attempt to shoot and kill some innocent person, but your shot is deflected by a falling rock whereas mine succeeds, according to one strand of common-sense thought we are equally blameworthy. But it is also undeniable that our reactions in reality often do depend on luck. So of two people who voluntarily get drunk and drive home, the one who kills a child will be blamed considerably more severely than the one who does not. If anything, here Sidgwick missed an opportunity to point out a further tension in common-sense morality, noted over a century before by Adam Smith (1976: 2.3.1; see also Williams 1981a; Nagel 1976).

Sidgwick frames his discussion of voluntary action as if he is attempting to describe a conception of it common to all ethical theories. It is more plausible to see it as the expression of his own view. By the 'reflective persons' who refuse to blame negligence, he means those who agree with him. And his restriction of attributions of rightness and wrongness to volitions or voluntary decisions alone depends on understanding those notions in their 'strictest' sense—a sense which he has himself given them. Further, as I have already implied above, it may seem surprising that this restrictive view of the scope of ethics should be recommended by a consequentialist, for whom anything can be an object of blame.¹¹ I suspect the explanation can be found in the nature of Sidgwick's rationalist project. As he sees ethics, it is primarily concerned with the question of what reasons we have to act. And if we can find an answer to that question which

11. It might be said that this is a result of Sidgwick's attempting to be neutral between consequentialism and deontology. But since the restrictive view suggests that consequentialism, as we now tend to understand it, is too broad, it is more likely that Sidgwick took the restrictive view of ethics first, with the (restricted version of) consequentialism coming later.

provides sufficient guidance for the living of a life, then we have no need to go into the issue of whether there are other objects of moral appraisal, such as one's character. But it cannot be denied that it is really far too early in the book for this conception of ethics to be working so hard, given Sidgwick's aim of discussing common-sense morality impartially. He may be able to persuade us that common sense is wrong to treat objects other than decisions to act as appropriate objects of moral appraisal; but any such attempt to persuade would come better after rather than before the analysis of common sense itself.

3.2 The Scope of the Will

After the completion of the main argument of 1.5, to the conclusion that the outcome of the free will debate is unimportant to ethics, Sidgwick returns in the final section of the chapter to the question of the power or scope of the will. This, he suggests, is an issue orthogonal to that of free will, since even a determinist must recognize the existence of volitions as psychological phenomena. And, he claims, this *is* an issue of significance for ethics, since ethical judgements, strictly understood, cannot apply beyond the boundary of what can be willed. As we have seen, Sidgwick exaggerates the role of volitions in ethics, but they are of course important, and the section contains some fascinating and characteristically incisive points in the philosophy of action.

Sidgwick is interested in the results of willing, and he finds them to fall 'mainly' into three classes: I: changes in the world beyond the agent, brought about through muscular contraction; II: changes in one's train of conscious thought or feeling; III: changes in tendencies to act.

One immediate question is whether Sidgwick has left anything out (note his 'mainly', above). I appears to exclude, for example, volitions *not* to act in some way or other, and these on one plausible understanding of causation can be said to have major effects. A volition not to increase one's donations to Oxfam, for example, may result in large amounts of unnecessary suffering. And II does not include the effects of willing on the subconscious or unconscious. But since Sidgwick's initial focus on volitions alone is anyway questionable, pointing out exceptions to his three classes at this stage is likely to be fruitless.

Sidgwick sees class I as the most prominent sphere of volitional causation. He begins by refuting the suggestion that it is the muscular

contraction itself that we will and not any more remote effect, since it may be that nothing follows from that contraction, pointing out that the muscular contraction itself may fail to result from a volition, as in cases of paralysis. Nor can we go back further into the causal chain, claiming that certain neural events are being willed. For we are usually quite unaware of these (and indeed of the muscular contractions). But Sidgwick does accept the idea that, nearly always, our will affects the external world through muscular contraction, and that our own causal role in bringing about these effects is complete. So, to use Donald Davidson's famous example (1963: 686), my role in the causal account of the light's being turned on ends with those muscular contractions that cause the switch to move.

Here we see that Sidgwick's suggestion that we limit the scope of ethical judgements to the limits of the will cannot be upheld. He himself has already stated that the scope of such judgements concerns the willing of certain results, such as, say, the switching on of a light or the exposure of a burglar, or the consequences of such exposure. Further, the muscular contractions are not themselves consciously willed, in most cases. So that would leave ethical judgements with nothing, in most cases, to be about. In other words, Sidgwick should separate the question of the causal role of willing from that of the appropriate object of moral judgement.

His position on causality here—that agential causation stops, in effect, within the agent's body—has proven attractive, in various forms, to many philosophers of action, especially in attempts to describe 'basic actions' (Danto 1963). But others will claim that the approach is reductive. Actions can be described in many ways—as muscular contractions, switchings, exposures—and some of these involve reference beyond the body of the agent. Or the whole approach of volitionalism might be questioned: what is being described as itself an act of willing may be said to be not a separate act but, say, a moving carried out in a particular way; alternatively, it may be objected that the term 'volition' in fact refers to a number of discrete mental acts. Debate continues, and we can conclude only that Sidgwick's view is philosophically respectable, yet no less controversial than any other view on the nature of action. As we shall see, this is perhaps another place in which Sidgwick might have gained from conceptual minimalism. What matters ethically, according to him, is that the world go as well as possible, and agents should play whatever role is required of them to bring about that best possible history. Whether that history is a result of their volitions or not is, from

the ethical point of view as he understands it, irrelevant, however interesting it may be as a metaphysical question.

Let us move on to category II. Can we control our feelings? Sidgwick shows how we can do so, at least indirectly. Imagine that I am very angry at something you have done and I am about to shout at you. I decide to resist my anger, however, and this may well decrease it, since feelings tend to be sustained through their expression. We cannot control our thoughts in the same way because they do not always prompt us to specific courses of action. But we can, to an extent, direct our trains of thought in particular directions, and ‘free ourselves of many thoughts and feelings upon which we do not wish to dwell’ (74). Here Sidgwick shows awareness of some of the basic psychological mechanisms underlying cognitive behaviour therapy, which would now almost certainly play a major role in putting any project of universal hedonism into practice (see Westbrook et al. 2011).

Sidgwick pays most attention in this section to category III changes brought about by volitions, since ‘it is in making general resolutions for future conduct that it is of most practical importance for us to know what is within the power of the will’ (74). Sidgwick claims that this question is at times confused with that concerning free will. Sometimes it is thought that libertarianism requires us to hold the view—later developed by Sartre as a component of his existentialism—that we are free to alter our habits at any time. Sidgwick suggests that the evidence of experience is against this (which Sartre would no doubt have described as ‘bad faith’ on Sidgwick’s part), and notes that any account of commitment on the basis of such libertarianism will be in internal tension: if I am free to bind myself to act in some specific way at some future time, then when that time comes I will no longer be free. But he claims that, since such commitments often do succeed, and even when they fail often do so only after a struggle, it is reasonable to impute to any commitment some motivational force. Sidgwick suggests that this force might consist either in a new, rational motive, or in a weakening of the power of habit. This distinction is an excellent example of the acuity of Sidgwick’s psychological insight.¹² Consider also his case of someone in the habit of drinking too much brandy during the evening, who commits himself to giving it up. In one scenario, his reason struggles

12. See also 75 n, in which Sidgwick provides an argument against an attempt to collapse category III effects of volitions into category II, distinguishing between cases in which a volition directs our attention to the benefits of breaking a habit (category II) and those in which we make a genuine resolution.

with his bodily appetite, and the thought that he has made a commitment he wishes to keep may assist his reason; in another, the force of habit has been more directly weakened by the commitment. Sidgwick concludes by pointing out that, since such willed commitments do sometimes succeed, determinists have as little justification for excessive pessimism about them as libertarians for excessive optimism: ‘On neither theory is it reasonable that we should deceive ourselves as to the extent of our weakness, or ignore it in the forecast of our conduct, or suppose it more easily remediable than it really is’ (76).

4 Determinism and Libertarianism

4.1 Determinism

In addition to intention, Sidgwick’s conception of volitions (that is, decisions to act or *choices*) includes ‘the consciousness of self *as choosing*, resolving, determining’ the results of action (61; my italics). Of course, it is not as if a volition consists only in intention (the representation of consequences), and the consciousness of movement and of choice; these can be understood in terms of their being properties of volition or choice itself. With his account of voluntary action in place, Sidgwick again states what he sees as the main issue in the free will debate:

Is the self to which I refer my deliberate volitions a self of strictly determinate moral qualities, a definite character partly inherited, partly formed by my past actions and feelings, and by any physical influences that it may have unconsciously received; so that my voluntary action, for good or for evil, is at any moment completely caused by the determinate qualities of this character, together with my circumstances, or the external influences acting on me at the moment—including under this latter term my present bodily conditions?—or is there always a possibility of my choosing to act in the manner that I now judge to be reasonable and right, whatever my previous actions and experiences may have been?

We may assume that Sidgwick’s determinist would claim also that any past actions of mine which led to my having the character I now possess would themselves have been determined by causes over which the agent had no control. Sidgwick goes on (62), plausibly enough, to deny that determinism must be materialist, and to expand the scope of the debate beyond moral choices: ‘The substantial dispute relates to the completeness of the causal

dependence of *any* volition upon the state of things at the preceding instant' (my italics).

Sidgwick then proceeds to offer a 'cumulative argument of great force' for determinism (62–5). He claims first that human enquiry has reached the point at which all 'competent thinkers' believe that all events, except human volitions, are causally determined. Science both depends on the assumption, and confirms it. We might expect, then, that human action will also be explained in the same way, especially given the 'increasing conviction of the essential unity of the cognizable universe'.

Sidgwick died just before the development by Max Planck and others of quantum theory, according to which, on most interpretations, matter behaves indeterministically at the most basic level. Of course, quantum theory itself, as well as its implications for the possibility of free will, is still much in dispute. These developments have weakened Sidgwick's first argument for determinism. But many continue to believe in determination at higher levels, and one might anyway make the position disjunctive: actions are either determined or random. And then Sidgwick's argument goes through.

The second argument—or set of considerations—Sidgwick offers for determinism focuses on the kind of unconscious actions he has already characterized as involuntary (63). First, we are asked to accept that unconscious actions are physically determined. Then Sidgwick suggests that the line between unconscious and conscious actions is vague, that conscious actions often become unconscious through habit, and indeed that any kind of conscious action could, in certain circumstances, have an unconscious origin. Finally, he returns to the 'impulsive actions' distinguished earlier from those resulting from deliberation (61),¹³ and claims that they lack the 'consciousness of choice' characteristic of deliberate actions to the point that we have no reason not to think of them as determined, and—again—that the line between impulsive and deliberate actions is unclear.

These considerations are unlikely to impress a libertarian. That some boundary is vague does not show—especially in the absence of anything like a sorites argument—that there is no boundary. We cannot specify any exact speed at which driving slowly becomes driving moderately, but this is not to deny that driving at 2 mph in a built-up area is driving slowly, while driving at 18 mph is driving moderately. The libertarian does not have to

13. These include the 'instinctive' actions of 61.

question the claim—though of course it is highly questionable—that any kind of action could become unconscious. The question is whether actions which are not unconscious are freely chosen. Nor need she dispute the claim that impulsive actions lack the ‘phenomenology of freedom’, since the implicit argument in which this claim is embedded itself depends on the same error about vagueness we saw above.

Sidgwick’s final argument suggests that determinism is implicit in our explanations of others than ourselves (63–5). Social life depends on our being able to make precise predictions of the actions of others, collectively or individually, and if some prediction fails, then we put it down to our lack of knowledge of character, not to the exercise of free will. Further, we even take this perspective on our own actions, once they are in the past or when they are in prospect. At the time we act, we have the ‘consciousness of choice’. But afterwards we will explain our action as ‘an effect of our nature, education, and circumstances’. And, when it comes to actions in the future, we believe that we can improve ourselves by determining them. Of course, sometimes we do believe the libertarian view that we shall in future be able to resist temptations to which we have always surrendered in the past. But this belief, as all moralists will agree, is largely ‘illusory and misleading’.

Again, this argument on its own cannot be said to possess the ‘great force’ Sidgwick attributes to his cumulative argument for determinism as a whole. There is first an internal tension in his own statement of it. In a footnote to the first sentence of 63, according to which we ‘always explain’ the voluntary action of others in deterministic terms, Sidgwick suggests that when we are *judging* the conduct of others morally we assume that they have free will, but that we do not ordinarily see it as a cause (and so as part of the explanation of the action). But free will, as Sidgwick understands it, is contra-causal, so in the case of any moral decision, the choice itself cannot be determined. Further, the libertarian is bound to seize on this admission by Sidgwick to suggest that, at least in the case of moral actions, we find an exception to the claim of universal causal determination. The same tension emerges in the idea that we take different temporal perspectives on our actions. If I can remember the consciousness of choice when I acted, why should I later view my past actions as determined? And if I know that the same consciousness will occur in my future actions, why should I now take that ‘consciousness’ to be an illusion? Further, when it comes to prediction, the libertarian may accept that people *usually* act according to character, but this in itself does nothing to show that free choice is impossible.

The arguments of 63–5, then, are perhaps best forgotten. But, as I suggested, Sidgwick's first, causal, argument at 62–3 might be restated in a powerful form. And his general background position, that there are impressive arguments on both sides of the debate, is hard to deny.

4.2 Libertarianism

As we have seen, the central claim of libertarianism, as Sidgwick construes it, concerns moral choice. It suggests that someone who voluntarily decides to perform some prohibited action not only could have acted rightly had she so chosen, but could have so chosen (59; 61–2). In this sense, then, libertarianism is more ambitious than compatibilism, which allows an action to be free merely if the agent could have acted otherwise had she chosen otherwise—even if in fact she could not have chosen otherwise.

The only argument Sidgwick offers for libertarianism is from the phenomenology of freedom: 'the immediate affirmation of consciousness in the moment of deliberate action' (65). In fact, however, Sidgwick fails fully to describe the content of the experience which the libertarian argument requires, saying:

Certainly when I have a distinct consciousness of choosing between alternatives of conduct, one of which I conceive as right or reasonable, I find it impossible not to think that I can now choose to do what I so conceive,—supposing that there is no obstacle to my doing it other than the condition of my desires and voluntary habits,—however strong may be my inclination to act unreasonably, and however uniformly I may have yielded to such inclinations in the past.

A determinist may allow that human beings sometimes have a motivation to duty of such strength that it will override any countervailing motivations. What matters for the libertarian is not this overriding motivation, but the notion of choosing itself.¹⁴ That is, in ordinary cases of decision,¹⁵ the future seems, as far as that decision is concerned, to be, to some extent, open: that is what 'choosing' is. As Sidgwick puts it earlier in the chapter: 'In deliberate volitions there is always a conscious selection of the result as one of two or more practical alternatives' (61). In other words, we have to assume that Sidgwick has in mind not only the consciousness of an overriding moral

14. That Sidgwick recognizes this is clear from 65 n, where he also notes that mere randomness is insufficient for libertarian freedom.

15. Such as those not involving other agents who will prevent my acting otherwise; see Frankfurt 1969: 835–6.

motivation, but also a sense of genuine choice. He feels that he can choose either to perform the right action or to go along with his non-moral desires.

As was also true in his presentation of the cumulative argument for determinism, Sidgwick is not seeking to persuade us of libertarianism. What he is attempting is to set out the position, and to show that it should be taken seriously. It seems to me that he has done that. There are nevertheless some serious difficulties with the argument from phenomenology. The most obvious one is the possibility that our experience of freedom could be illusory, an implication of Sidgwick's own causal argument for determinism. There is also some challenging psychological evidence against libertarianism (e.g. Libet et al. 1983). Nevertheless, many remain unshaken in their belief that our 'consciousness of choice' is genuine.

Sidgwick does indeed consider the possibility that he is under an illusion, imagining his seeing an action against his rational judgement as determined (65–6). He continues:

But I cannot conceive myself seeing this, without at the same time conceiving my whole conception of what I now call 'my' action fundamentally altered: I cannot conceive that if I contemplated the actions of my organism in this light I should refer them to my 'self'—*i.e.* to the mind so contemplating—in the sense in which I now refer them.

In the light of Sidgwick's own version of compatibilism, according to which the free will debate is practically irrelevant, this passage might initially appear rather mysterious. But note that he does not claim here that he would no longer see his action as free. Rather, he would no longer attribute it to his self in the way in which he usually does. What is usually the case is that I do not *know* which of two actions I am in fact going to perform, so I will engage in some process of deliberation which will issue in the choice of a particular action. In the kind of case Sidgwick is imagining, there would be no need for deliberation, and so the activity of 'the self' would play no part in explaining the action. Sidgwick nowhere says that his accepting determinism would lead to his having to revise his conception of action and the role of the self in action. He speaks here only of those unusual cases in which I know what I am going to do.

5 The Significance of the Debate

Had Sidgwick attempted to sidestep the debate about free will through an appeal to Humean compatibilism, he would have failed to persuade

incompatibilists, whether determinist or libertarian. So in sections 3 and 4 of 1.5, as we have seen (CD 2.2), he seeks to show that whichever position turns out to be correct has no implications for ethics.

Sidgwick suggests that it is libertarians in particular who have exaggerated the importance of the debate, and he begins by outlining their argument (66). Essentially, their claim is that, since we can properly hold a person to account morally if and only if they could have acted otherwise, the determinist denial that she could have acted otherwise—since her actions are the results of chains of causation stretching back beyond her very existence—is inconsistent with such holding to account.

Sidgwick's first move is to reintroduce the compatibilist conception of 'acting otherwise' as what often underlies the notion that people can be morally bound to do only what is in their power (67). But he is quite aware that the libertarian will insist that I should be free *to choose* to act rightly, not just to do what I have chosen. He again states his view that he 'inevitably conceive[s]' that he can so choose, admitting the phenomenology without committing himself to libertarianism itself, but once more imagines his supposing that phenomenology to be an illusion and judging on the basis of past performance that he will certainly not choose to do what is right. Previously, Sidgwick spoke of the effects of such a supposition on his conception of action and self. Now he admits that it will weaken and perhaps entirely undermine any motivation to act morally. Either he will not judge it reasonable to choose to do what he would otherwise judge it reasonable to choose, or, if he does so judge, he will also judge the notion of duty embedded in that judgement to be as illusory as the conception of freedom.

This might appear to be an admission of the significance for ethics of the free will debate. But it is not. Sidgwick concludes this paragraph (68) with the claim that the debate matters only if it affects 'my view of what it would be reasonable to choose to do if I could so choose'. In the case he has imagined, of course, that view is affected. But, as we have seen, he believes such cases to be extremely rare.¹⁶ Usually, that is to say, the future is unknown to me. It may be that my consciousness of freedom is veridical, and I make genuine choices; or it may be an illusion, but in that case it still makes sense

16. At 68 n. 1, Sidgwick notes that there may be cases in which my knowledge that in the future I will probably not will to act rightly may be a reason for my not willing now to do what would otherwise be right, if the rightness of my willing now depends on what I will in the future. See Jackson and Pargetter 1986.

for me to 'will' to do what is right and reasonable—because it could be that this is what I am determined to do.

We are now in a position to see how Sidgwick's consequentialism may have affected his conception of practical reasoning and its components. According to him, what matters morally are decisions, and those decisions are to be judged by their results. So, if I am aiming to do what is right, I must try to will those results that are best. If I know that I will not and hence cannot will the results that would be best, then, as Sidgwick points out, it is unlikely that I will try. But if there is even a small possibility of my willing the best results, then, Sidgwick believes, it would be unreasonable of me not to try—since what I am trying to will is itself what I take to be most reasonable in the circumstances. So, for Sidgwick, the significance of the debate between libertarianism and determinism is to be judged solely by its implications for the prediction of the results of willing. But most libertarians, especially if non-consequentialist, are likely to think that what it is reasonable for me to will depends upon assessments of the results of my will that depend heavily upon concepts that themselves make sense only within a non-deterministic framework (and of course incompatibilist determinists will agree that these concepts require non-determinism). As we shall now see, Sidgwick did recognize that these concepts were threatened by determinism. But his solution is to provide consequentialist interpretations of them which most libertarians will reject outright.

At 68, Sidgwick asks whether the outcome of the debate might affect our view of the ends we believe it rational to pursue. He has already claimed that those ends are happiness, and perfection or excellence (6–11). Sidgwick sees no reason for a libertarian who accepts happiness as a rational end to reject it, should she be converted to determinism. Such a person may well believe that the greatest balance of pleasure over pain is good for a certain individual, or just good, if the pleasures and pains are those of all sentient beings. But it is not clear that she must be prepared to accept that the idea of a 'rational end' is consistent with a deterministic framework. An end, she may claim, must be something that we can freely choose; if there is no genuine choice, the results of our 'actions' are not anything that we can properly be said to *aim* at. Our behaviour may be directed at the good, but we are not directing it. Even if she is ready to accept that rational ends may be determined, she may deny that happiness itself is unrelated to the outcome of the free will debate. It may be rational to continue to aim at certain pleasures and avoid certain pains—the physical variety, for example. But imagine that

our libertarian had taken much pleasure in her accomplishments and her moral integrity, attributing both to free and unconstrained choices. Especially in the case of moral integrity, it is not unlikely that belief in determinism may affect these more sophisticated, intentional pleasures, and indeed cause her a certain amount of unpleasant regret.

A similar problem arises in the case of excellence. Sidgwick suggests that physical and intellectual perfection are both independent of free will. Seen merely as perfections, this may be true. But if either is seen as a personal accomplishment, then the libertarian may see determinism as undermining their value. Sidgwick goes on to make a similar claim about the moral excellences: the exercise of the virtue of, say, temperance is no 'less admirable because we can trace [its] antecedents in a happy balance of inherited dispositions developed by a careful education'.¹⁷ The obvious worry about this is that the admiration we feel for such actions is admiration for the choice or decision made by the agent, and hence for the agent as the kind of person to make such choices. Indeed, Sidgwick accepts this as a worry, noting (68 n. 2) that determinism cannot make room for the 'ordinary notion of merit' (see 349).

One might be tempted to wonder at this point why Sidgwick does not claim that this is in fact a pseudo-worry, since the idea of merit or desert requires only compatibilist, not libertarian, freedom. But of course his aim is to show that it makes no practical difference whether one is an incompatibilist libertarian or a determinist. His response to the worry, then, is to claim that our admiration in such cases does not require that that admiration be merited. We can see this in the case of God, whose nature we can see as perfect and hence as admirable, even though the notion of merit is inapplicable to him. The libertarian might respond, however, that the reason why merit is inapplicable to God is his perfection: it is part of his essence always to act rightly and rationally, and so the kind of moral praise reserved for those who do what they 'ought' to do is inappropriate in his case. That is not true of imperfect moral agents like us.

Having discussed the significance of the debate for rational ends, Sidgwick turns to the question of the *means* to those ends (69–70). Sidgwick begins with the position that the world is morally governed by God, so that performing one's duty is the best means to happiness because God will reward

17. Note also the discussion of courage as an excellence rather than a virtue, and so 'not completely voluntary', at 333–4. It is still, according to Sidgwick, a subject of '*moral estimation*' and '*morally admirable*' (my italics).

one in the afterlife. And he allows that this position would be seriously weakened by determinism: 'This is the main element of truth in the view that the denial of Free Will removes motives to the performance of Duty.' Sidgwick rules out as irrelevant to his project any discussion of the theological argument itself, but in fact what he says a little later about the notion of desert suggests one position he might take on it. At 71–2, Sidgwick suggests that the 'only suitable meaning' for desert is in terms of the effects of punishment and reward. It is true that this is the only suitable meaning 'from an utilitarian point of view'; but that point of view is Sidgwick's own, and has the implication that the theological argument discussed in 1.5.3 attributes to God an incorrect understanding of merit and desert (since deterrence would operate here not through God's actually punishing us in the afterlife but through our believing that he would). Indeed, the weight Sidgwick attaches to this argument is odd. Just like God, we attribute rewards and punishments in this life on the basis of how morally well or badly we believe someone to have acted. These attributions could also be said by the libertarian to rest on the assumption of free will.

Sidgwick then turns to those means to ends which are 'scientifically cognisable' (69–70), and considers whether the outcome of the free will debate is relevant to predictions about the future based on what has happened in the past. Sidgwick claims that, even if we admit free will as a cause which might falsify scientific predictions of action, we will not be led to modify those forecasts, since as a cause free will is 'absolutely unknown'. He illustrates this argument with a bizarre example. Imagine that we thought the planets had free will and maintained their motion only through exercising it continually. Our confidence in the future of the solar system, Sidgwick suggests, would indeed be impaired; but our astronomical calculations would be unaffected. The thought here is that all the astronomer can do is make predictions based on the past. Ordinarily, once she has finished her calculations, if she believes she has done them properly, then she will probably be quite happy to say that she knows for sure that, say, Mars will be at some specific point in this evening's night sky. On the assumption that Mars has free will, however, she should be less sure; for Mars might 'decide' to act differently from the way it has acted in the past. And this may have practical implications. Imagine, for example, that some fellow astronomer has asked her to bet on where Mars will be tonight. This is indeed what one would expect from a contra-causal conception of freedom. So, just as with ends, the outcome of the free will debate is also relevant to our judgements

about means. Sidgwick appears to ignore the importance of probability-judgements in practical reasoning. The assumption of contra-causal free will must affect the probabilities attached to judgements based merely on past behaviour.

Sidgwick then (71–2) returns to the potential inconsistency between determinism and central moral concepts. He begins with remorse, and, again, rather than attempting a compatibilist account of the concept, he admits that ‘so far as the sentiment of remorse implies self-blame irremovably fixed on the self blamed, it must tend to vanish from the mind of a convinced Determinist’. Further, as Sidgwick accepts, this may affect the *motive* to right action, even if the reasons for it remain unaltered by the assumption of determinism. In fact, it is hard to see why the effect on the motive is not itself an effect on the reasons: one of my reasons for refraining from immoral action is the remorse that I shall feel afterwards. To the extent that I believe that I will not feel that remorse, that reason itself is diminished.

Sidgwick’s response to this inconsistency is essentially to weaken considerably the central claim of his chapter (see Fowler and Selby-Bigge 1890: 95). In the end, it amounts to the claim not that the free will debate *is* irrelevant to ethics, but that it *may* be irrelevant. If one is a determinist with imagination, sympathy, and love of goodness, it may be that the place of remorse in self-improvement will be taken, as Sidgwick suggests, by dislike of those aspects of one’s character that have led to immoral action in the past. (The expectation of such dislike—given that the dislike itself is likely to be unpleasant—may also provide an *ex ante* reason in favour of moral action.) Although Sidgwick does not go on to state it, his view appears to be that if one lacks sufficient imagination, sympathy, and love of goodness, then it will make a difference to one’s practical reasoning and action, whether one is a libertarian or a determinist.

Sidgwick then considers punishment more generally (reward is also mentioned), and again admits that our ordinary conceptions of responsibility, retribution, and merit depend on the assumption of free will:

If the wrong act, and the bad qualities of character manifested in it, are conceived as the necessary effects of causes antecedent or external to the existence of the agent, the moral responsibility—in the ordinary sense—for the mischief caused by them can no longer rest on him. (71; see 284–5; 291)

Once again, Sidgwick’s response is to refer to a ‘forward-looking’ alternative justification for punishment, based on deterrence and encouragement,

and he suggests that the notion of 'responsibility' underlying such practices is the 'only suitable' one, 'from an utilitarian point of view'. In other words, *if you are a utilitarian, the outcome of the free will debate will be irrelevant, since your conception of responsibility and related notions does not, unlike the common-sense conception, rely on the assumption of free will.* The unstated implication of this, of course, is that if you are not a utilitarian, then the outcome of the free will debate may matter a great deal.

Sidgwick clearly recognizes this danger to his argument, and refers us to 3.5, where he claims he will show that retributive and utilitarian theories, though quite different, have very similar practical implications (285). For example, we will need anyway to punish people for reasons of social order, regardless of retributive considerations, just as we punish people for negligence even when we cannot trace it to a decision to ignore one's duty (72 n. 2). Further, the retributive view of punishment clashes with benevolence, since it advocates punishment 'as a purely useless evil'. Sidgwick's final point mirrors the point he made about remorse. A determinist will no longer wish to praise and blame others morally on grounds of desert. But she may encourage good and discourage bad conduct in her own way, and as effectively as a libertarian.

Sidgwick is here somewhat complacent about the effects of utilitarian thinking on our moral and judicial practices. As he admits, retributive punishment in itself is a 'useless evil', and it is a question at least worth discussing whether utilitarian practices of deterrence and reform might not be considerably more humane than the system of prisons and other punishments standard in Sidgwick's and our own time. And he is complacent too about the utilitarian value of retributive notions themselves. Guilt, blame, shame, and positive analogues such as moral satisfaction, a sense of integrity or decency, and moral pride have their own internal logic, and are, though mistaken according to utilitarianism, on the face of it potentially very useful.¹⁸ It is not clear why Sidgwick himself fails to make this point, since he will go on later in the book to see common-sense morality as itself justifiable only through its resting on a utilitarian foundation. Indeed, had he seen the point, as far as his main project in the *Methods* is concerned, he could have omitted his chapter on free will entirely (publishing it as a separate paper, perhaps), and included a footnote to his chapter on utilitarianism and

18. As Slote (1990: 382–3 n. 1) notes, a utilitarian might argue that both utilitarianism and (hard) determinism should be 'esoteric'.

common-sense morality (4.3) on the unimportance to ethics—other than in highly unusual cases—of the outcome of the free will debate.¹⁹ Of course, as I observed above in connection with systems of punishment, it may be argued that the implications of utilitarianism for that morality are significantly more radical than Sidgwick allows. But given his lack of radicalism, and the fact that his key arguments in 1.5 themselves depend on utilitarianism, a direct way is open to the conclusion Sidgwick becomes so entangled in arguing for in 1.5.

19. Alexander Bain (1876: 181) calls 1.5 a 'half-needful digression'.

3

Hedonism and the Ultimate Good

I Goodness, Rightness, and the Ultimate Good

Sidgwick is led into his first substantive discussion of the notion of goodness through noting a difference between intuitionists about the *object* of intuition (103). Up to this point, he has been speaking of that object as *rightness*, any attribution of which is taken to imply a dictate of reason to perform the action in question, either as right in itself or with a view to some further end. But in the first section of 1.9 (the chapter entitled just ‘Good’), Sidgwick contrasts rightness with goodness (see 2–4; 228; 391–2).

Sidgwick is, then, working with a normative conception of rightness. The question cannot arise for me whether, once I see some action as right, I have any reason to perform it. In the case of goodness, however, no ‘definite precept’ (113) need be mentioned, and the moral ideal here is to be construed as *attractive* rather than *imperative*. The distinction Sidgwick is drawing here, however, might seem more a matter of stipulation than of substance. First, we can make sense of a conception of rightness which is not analytically normative, such that an egoist, for example, might admit that some action is right, but doubt whether she has any reason to perform it except in so far as it promotes her own good or well-being. Second, there is nothing to prevent our stipulating a sense of goodness which is itself non-latently normative. Indeed, this sense follows from so-called ‘buck-passing’ accounts of goodness, according to which the goodness of any object is to be understood in terms of its possessing some (usually, natural) property or properties which ground certain reasons.¹ And at 112, Sidgwick himself says

1. See e.g. Scanlon 1998: 95–100. Note that Sidgwick himself was attracted to a buck-passing or ‘fitting-attitude’ account of goodness: see later in this paragraph.

it is ‘more in accordance with common sense’ to see the desire for one’s own good as involving a ‘rational dictate to aim at this end’. Nevertheless, if we take Sidgwick’s notion of rightness as intended primarily to capture the notion of normative reasons itself, there does—contra the buck-passing view—seem to be a strong case for distinguishing reasons from goodness. Normative judgements and evaluative judgements, that is to say, should be kept distinct from one another. It does not follow from my having reason to φ that φ -ing would be good in any non-trivial sense, nor from something’s being good that it grounds any reason.

Sidgwick’s understanding of the rightness/goodness distinction exaggerates the difference between ancient ‘attractive’ and modern ‘imperative’ ethics (see CD, chapter 1 n. 34; White 1992: 313). The notion of the good certainly plays a significant role in ancient ethics, but so does that of what we have reason or ‘ought’ to do. Consider, for example, Aristotle’s ‘doctrine of the mean’, according to which the aim of a virtuous person will be doing and feeling what she *ought* (*dei*) (1894: 1106b21–3). The reason the ancients focused on the question of the relation of virtue to the agent’s own good is not, as Sidgwick suggests, that they were using a generic notion—‘good’—and so were led to examine the relation of two different species of goodness. Rather, like modern thinkers, they were struck by the force of the egoistic challenge to morality: why should I do what is right (that is, morally good or virtuous) if it does not advance my own good?² That question itself may involve the non-normative conception of rightness that Sidgwick fails to recognize, and this perhaps goes some way to explaining why he gave an account of its origin among the ancients solely by reference to the notion of goodness.

Having dealt with ‘synonymic hedonism’ (see CD 3.3), Sidgwick raises, with a quotation from Hobbes, the question of whether the meaning of the term ‘good’ can be given in terms of what a person desires (109–10).³ The tautology argument against synonymic hedonism (109) would also work against any such proposal. But in fact Sidgwick’s discussion in this section is best understood as about not the meaning of words, but what the good is. Is a person’s own good—what is good for her—to be understood as her obtaining what she desires non-instrumentally for herself?

2. By ‘modern’, I mean ‘non-ancient’. Perhaps partly because of the influence of Moore’s (1903: 97–9) attempt to undermine egoism, the question of how to justify morality to the normative egoist was less prominent in twentieth-century ethics than in previous centuries. See CD 7.1.1.

3. As noted by de Lazari-Radek and Singer (2014: 201 n. 1), the quotation is slightly inaccurate.

One obvious objection is that people often desire what they know is overall bad for them. Sidgwick suggests that we distinguish between the desired 'result' (e.g. the taste of champagne) and its consequences (e.g. severe indigestion), and allow that present desire does not always properly respond to the representation of bad consequences. Nevertheless, the badness of the consequences can be understood in terms of the later and stronger desires for their absence. Having restricted the theory to desired results alone, independently of their consequences, Sidgwick then makes the powerful point that what is desired can be only the *apparent* good, since the objects of our desire may turn out to be valueless, or less good than we expect. (They are also of course sometimes better than we expect.) Further, prudent people will often try to suppress desires for things out of their control, such as good weather, and yet will not thereby see the objects in question as any less good.

These worries lead Sidgwick to claim that we should focus on the *desirable* rather than the *desired*. But his understanding of desirability here is unusual. He means not (or 'not necessarily', as he puts it) what *ought* to be desired, but:

what would be desired, with strength proportioned to the degree of desirability, if it were judged attainable by voluntary action, supposing the desirer to possess a perfect forecast, emotional as well as intellectual, of the state of attainment or fruition. (111)

One immediate oddity is the inclusion of the explanandum—desirability—within the explanans. Perhaps what Sidgwick means here is that, on this account, the object of any desire should be seen as desirable to the extent that it is desired. So if I desire some object twice as much as another, then it is twice as desirable. The restriction to objects that can be achieved through voluntary action is, perhaps, an attempt by Sidgwick to take into account the point in the previous paragraph about the prudent suppression of desires for objects out of one's control. But the reason the prudent person does that is to minimize the pain of disappointment. When she is considering certain *hypothetical* outcomes, she has no such reason, and, if she foresees accurately the effects of good weather, can rationally desire that outcome just as strongly as some outcome within her control. Finally, we should note that it would have been more accurate had Sidgwick spoken of a move here from actual to hypothetical desire, not from the desired to the desirable. 'Desirable' can mean 'ought to be desired'; perhaps it can mean 'can be desired'. But it certainly does not mean 'would be desired'.

Sidgwick again notes the problem of ‘bad consequences’, and suggests that this problem could be dealt with if we restrict ourselves to ‘desire which becomes practical in volition’—that is, desire that I would will myself to act on. He must also be assuming a lack of weakness of will, since in the champagne example I would will the drinking of the champagne even in the knowledge that it would be overall worse for me. This leads us to the following account:

my Good on the whole is what I should actually desire and seek if all the consequences of seeking it could be foreknown and adequately realised by me in imagination at the time of making my choice. (111)

Sidgwick approves of the temporal neutrality implicit in this account, but does draw attention to a serious problem: it does not sufficiently incorporate comparison between options. Consider a case in which I am considering some course of life *L*, which I know will include a large number of pleasures, few pains, and no sense of regret. I cannot assume that my good consists in *L*, however, since lack of regret is not proof of one’s having made the right choice. So what we need is:

a man’s future good on the whole is what he would now desire and seek on the whole if all the consequences of all the different lines of conduct open to him were accurately foreseen and adequately realised in imagination at the present point of time. (111–12)

Sidgwick is considerably happier with this account than those leading up to it, noting that, though it contains an ‘ideal element’, since not everyone does actually desire the good so understood, this element itself can be understood entirely in terms of non-evaluative *fact* rather than value or any dictate of reason. This is an advantage, one presumes, because the account is more likely to be acceptable to those who wish ethics to be naturalistic.

Sidgwick himself, however, is not a naturalist of this kind, and prefers to see—along with common sense and Butler—the desire for my good on the whole as *authoritative*, involving an implicit rational dictate to aim at this end in the face of conflicting desires. But we can keep that reference ‘latent’ by understanding ultimate good as ‘what I should practically desire if my desires were in harmony with reason, assuming my own existence alone to be considered’ (112).⁴ I understand Sidgwick to be suggesting the incorporation

4. Schneewind (1977: 224) plausibly takes ‘should’ here as equivalent to ‘would’; see also Frankena (2000: 283).

of the notion of harmony into the earlier definition, rather than offering an independent alternative, though it remains unclear why he does not add a clause to that definition.⁵

The rationalist account does avoid any problems arising for the earlier definition from the possibility of ‘deviant’ desires, such as that in Rawls’s case of the person who, having foreseen all the various options open to her, desires most strongly a life of counting blades of grass (1999: 379–80). But the view is a version of what Rabinowicz and Rønnow-Rasmussen call a ‘fitting-attitude’ (FA-)analysis of goodness, and it is in danger of running into what they call ‘the wrong kind of reasons objection’ (2004: 393). Consider a case in which an evil demon threatens to inflict great agony on me if I do not desire something, such as the possession of a saucer of mud, that appears to be irrelevant to what is ultimately good for me (Crisp 2000: 459). I now have a strong reason to desire the mud, but it is hard to see why we should accept its possession as a constituent of my welfare or well-being. This objection is still controversial, and several lines of response have been offered to it on behalf of FA-analyses. But what lies behind the objection is the dependency of desire upon goodness, which any account of goodness in terms of desire is likely to reverse. We desire things, such as pleasure, *because* they are good, and often because they are good *for* us. ‘Good for’, then, is better seen as a primitive notion, one that should not, and need not, be elucidated in terms of fitting attitudes, whether desires or not. This is another place in the *Methods* where Sidgwick gets himself into unnecessary difficulties.⁶

5. Other interpreters appear to prefer the view that Sidgwick replaces the earlier definition with the later. See e.g. Broad 1930: 176; Frankena 1956: 541; Schneewind 1977: 224; Parfit 1984: 500; Shaver 1997: 314. (Frankena later changed his mind (2000: 282) on the basis of Sidgwick’s claim about implicitness and latency. His revised view seems to be that Sidgwick is suggesting that either definition is acceptable. But since the later one makes explicit what is implicit in the earlier one, I take my view to be close to Frankena’s revised position.) Shaver has suggested to me that it may be preferable to see Sidgwick as offering a new position rather than adding to the previous one because of his view that questions of goodness are to be settled by intuition (see 400). This suggestion is made also at M1 98, where Sidgwick speaks of ‘simple’ intuitions, whereas here he admits that the earlier statement is ‘elaborate and complex’. But he goes on to say that it is ‘intelligible and *admissible*’ (my italics). Nor do we have to think that whenever Sidgwick is elucidating some notion, he must always be appealing to an ultimate intuition. When it comes to what is good, the ultimate intuition is indeed that pleasure is the only good, not the claim involving hypothetical desire. Sidgwick perhaps sees a distinction between elucidating the idea of goodness and developing an account of what is good, only the latter of which must appeal to an ultimate intuition.
6. Sidgwick is also leaving himself open to the charge that we cannot assume that the ideally informed subject will desire what maximizes the balance of pleasure over pain in her life, so that his account of ultimate good is at risk of tension with his hedonism. Given his conception of the good as attractive rather than ‘non-latently’ imperative, I am inclined to think that, if pressed, he would have rejected any buck-passing account of goodness. For a different view, see Phillips (2011: 124).

Similar difficulties arise in the remainder of 1.9.3 (112–13). With the phrase ‘assuming my own existence alone to be considered’, Sidgwick has limited the account to the good of the subject herself. If we remove that, and include instead ‘an equal concern for *all* existence’, we have an account of the ultimate good on the whole, as opposed to the ultimate good for the individual. Again, however, this account runs into the objection that the content it gives to rationality is itself acquired from an independent notion of the good, so that the account itself is an idle wheel. On Sidgwick’s own view, if I am rational and impartial, and considering various possible outcomes, I will desire that which contains the greatest overall balance of pleasure over pain. But that is merely because, according to Sidgwick (understood as a non-buck-passer), that outcome is best. We are not helped to understand its being the best, or indeed its goodness, through this buck-passing account. The same goes for Sidgwick’s next claim—that this impartial notion of the ultimate good on the whole explains what it is for an action to be ‘good’ in itself, independently of its consequences. Someone who thinks that, say, telling the truth is good in itself may well accept that a rational and impartial subject would desire that such an action be performed. But, unless one wishes to expunge any substantive idea of goodness from one’s account (which in effect is what buck-passers are usually aiming to do, and which Sidgwick is most certainly not attempting to do), it is not clear what is added to our understanding by such an account of goodness.

Sidgwick concludes section 3 by returning to the distinction between rightness and goodness. Of the account of ultimate good, he says:

Such a judgment differs, as I have said, from the judgment that conduct is ‘right’, in so far as it does not involve a definite precept to perform it; since it still leaves it an open question whether this particular kind of good is the greatest good that we can under the circumstances obtain. (113)

It might seem that Sidgwick is beginning to distance himself from the contrast he drew between rightness and goodness, according to which only the former involves a ‘dictate of reason’. For could we not say that the judgement that some action is the *best* implies such a dictate? But Sidgwick goes on to argue that claims about the goodness of actions do not involve a commitment to the performability of the actions in question:

in fact there are many excellences of behaviour which we cannot attain by any effort of will, at least directly and at the moment: hence we often feel that

the recognition of goodness in the conduct of others does not carry with it a clear precept to do likewise, but rather

the vague desire
That stirs an imitative will.⁷

Sidgwick's point is that we cannot claim that attributions of ultimate goodness or bestness to actions imply a rational dictate to perform them because these actions may be impossible and hence such that they cannot be rationally required. Once we allow that rightness does involve a rational dictate, the idea that 'rightness implies can' seems plausible enough. But we might also want to claim that the idea of bestness implies a rational dictate to perform the action *if one can*. Further, the accounts of ultimate good Sidgwick has been discussing, explicitly to begin with and then later implicitly (as in the reference to what I should '*practically*' desire in the final account of the ultimate good for me), have been limited in scope to what it is possible for me to seek. With that restriction in place, there is no contrast here between rightness and goodness.

2 Pleasure

2.1 The Volitional View

Sidgwick interrupts his discussion of the nature of ethics and ethical judgements in 1.1–3 with a chapter on pleasure and desire. (The chapter on free will follows, before Sidgwick returns to the discussion of ethics in 1.6.) In the first section of 1.4, Sidgwick suggests that, since (a) there are forms of psychological hedonism which do not rule out ethical judgements and (b) psychological hedonism does indeed make egoistic hedonism seem *prima facie* reasonable, it is important to consider psychological hedonism. As part of that, he must of course consider the nature of the pleasure which the psychological hedonist suggests we seek.

Sidgwick begins by 'conceding' that 'pleasure is a kind of feeling which stimulates the will to actions tending to sustain or produce it...and similarly pain is a kind of feeling which stimulates to actions tending to remove or avert it' (42–3). The 'felt volitional stimulus' in each case Sidgwick calls *desire* and *aversion*, extending these notions, as he says, from cases in which pleasure

7. From Tennyson's *In Memoriam*, cx.

or pain is merely ‘represented in idea’ to those in which it is present.⁸ We might call this the *volitional view* of pleasure and pain.⁹

Since the volitional stimulus in question must be *felt*, an immediate question is whether Sidgwick is intending to identify that feeling (presumably along with the experience being willed to continue) with (the feeling of) pleasure itself. That is, is the feeling of pleasure *just* the experience of, say, eating or reading and the feeling of wanting to sustain or to produce that experience, or is the pleasure an independently identifiable feeling which brings with it another feeling—that of the felt desire to sustain or produce that feeling? The latter must be what Sidgwick has in mind, since the view involves a clear distinction between the desire to produce the feeling of pleasure and the feeling of pleasure itself. If I desire to produce the pleasurable experience of eating, for example, that will of course not immediately give me the pleasure of eating.¹⁰ Likewise, once I am enjoying eating, the felt volitional stimulus to sustain that experience must be distinguished from the enjoyment of eating itself. We might call this the *two-feelings* version of the volitional view, as opposed to the *identification* version, which would have to restrict itself to claims about sustaining rather than producing experience.

Why does Sidgwick say that he ‘concedes’ the volitional view? 43 n. 1 provides the answer: ‘The qualifications and limitations which this proposition requires, before it can be accepted as strictly true, do not seem to me important for the purpose of the present argument.’ That footnote refers us forward to 125–30 for further discussion. In 1.4 the main question is whether we seek pleasure (whatever it is); in 2.2 the issue is whether pleasure can be quantified, and so a more precise account is required.

8. In a footnote (43 n. 2), Sidgwick allows that desire can sometimes arise for objects known to be unattainable (see 113; CD 3.1). Here, the stimulus is not to action to attain the object in question but to ‘voluntary day-dreaming, by which a bitter-sweet imaginary satisfaction of the want is attained’. But this species of desire he sees as secondary.
9. Another name, one I myself used (2011a: 33), might be ‘preference hedonism’. I have chosen to avoid this because other versions of preference hedonism can be found in the literature, and there is also a danger of confusing views about what pleasure is with views about motivation (psychological hedonism) or value (evaluative hedonism).
10. Further support for this interpretation can be found at 43: ‘The question at issue, then, is not whether pleasure, present or represented, is normally accompanied by an impulse to prolong the actual or realise the represented feeling, and pain correspondingly by aversion: but whether there are no desires and aversions which have not pleasures and pains for their objects—no conscious impulses to produce or avert results other than the agent’s own feelings.’ I read Sidgwick here to be implying that the first question is answered affirmatively by those involved in the debate, and it involves a clear distinction between pleasure and the volitional impulse.

Sidgwick begins by attributing the volitional view to Spencer and Bain:

The equivalent phrase for Pleasure, according to Mr. Spencer, is ‘a feeling which we seek to bring into consciousness and retain there’; and similarly, Mr. Bain says that ‘pleasure and pain, in the actual or real experience, are to be held as identical with motive power’.

Unfortunately, Sidgwick does not here distinguish between the two-feelings and the identification versions of the volitional view. Spencer appears to hold the former (he goes on in the passage Sidgwick cites to speak of pleasures as ‘agreeable or desired feelings’), and Bain the latter (if we interpret Bain as holding that pleasure is a feeling, so as to avoid problems with unconscious motives).¹¹ If, however, we interpret the two-feelings view as the position that what *makes* a feeling a pleasure is the extent to which it gives rise to the felt volitional stimulus to sustain or produce it, Sidgwick’s central objection (125–7) to the volitional view—that there is no strict correlation between degree of pleasantness and strength of felt volitional stimulus—applies to both versions.

Sidgwick drops talk of ‘production’, which as I have said is difficult for the identification view to incorporate. But he does go on, in a spirit of charity, to elucidate the volitional view in two ways. The first point is that the volitional stimulus may not be to any particular exertion. So someone enjoying a warm bath after a long day’s work ‘is vaguely conscious of a strong clinging to his actual condition, and of a latent readiness to resist any impulse to change it’. Both Spencer and Bain might agree with this. The second point, however, is impossible for Bain to accept, since it suggests that ‘the stimulus of moderate pleasures and pains may become unfelt through habitual repression’. Sidgwick’s example of such a pleasure is that of a temperate person, in whom the stimulus to sustain the pleasure of eating usually ceases before the pleasure does so. Bain must insist that, as long as the experience of eating is enjoyed, the desire to sustain it must be present in proportion to the degree of enjoyment. For the feeling of having such a desire while eating is identical to the pleasure.

Sidgwick focuses on cases where the stimulus is strong, and finds an internal tension in Bain’s account. When speaking of the degree of pleasure and pain, Bain distinguishes between ‘intensity or acuteness’ and ‘quantity or mass’

11. In the passage preceding that cited by Sidgwick, Bain says: ‘the stronger motive in the shape of actual and present sensation or emotion *is* the greater pleasure, or the smaller pain’ (my italics) (1868: 356).

(1868: 217–18). The prick of a pin is acute, while a sense of fatigue is massive. Sidgwick quotes Bain as allowing that ‘acute pleasures and pains stimulate the will perhaps more strongly than an equivalent stimulation of the massive kind’, and could have added a passage which follows shortly upon that just quoted in which Bain allows that massive pains ‘debilitate and weaken the tone of the system, and are not favourable to voluntary exertion, although their motive force *ought* to be great’ (my italics). In effect, in these passages, Bain is giving up the identification for the two-feelings version of the volitional view, and then admitting that volitional stimuli do not correlate with degree of pleasure and pain. Sidgwick offers another characteristically apt example, noting that some feelings, such as that of being tickled, involve a strong aversive stimulus but are often not painful at all or only slightly painful.

2.2 The Feeling–Tone View

Wayne Sumner distinguishes accounts of pleasure as either *internalist* or *externalist* (1996: 87–91). According to internalism, an experience is pleasurable in so far as it has some ‘internal quality’. This is most commonly seen as a special kind of basic and unanalysable ‘feeling–tone’, and it is their possessing this feeling–tone that different pleasurable experiences have in common and explains our calling them ‘pleasures’. According to externalism, introspection reveals no such feeling–tone, and we have to understand pleasure in terms of an experience (such as eating) and some attitude towards that experience and hence ‘external’ to it (such as a desire for the experience to continue).

The discussion of the volitional view above, however, suggests that this way of drawing the contrast may be misleading. One might construct an externalist account according to which pleasure is constituted solely by the having of some experience (such as that of eating) along with some desire (such as that for the experience to continue). On one version of the view, the desire in question need not be *felt* for pleasure to come into being; it might even be an unconscious desire, and leave phenomenology unaffected. But this view (the *non-phenomenological view*) seems highly implausible, since it would be unable to distinguish between the phenomenology of mere eating on the one hand, and *enjoying* eating on the other. It may not be that the actual desire has to be felt. But the presence of the desire must at least make some difference to the experience—to how it feels—and that difference may be described with reference to the feeling of pleasure.

In other words, the difference between internalism and externalism is a difference in what kind of feeling pleasure is.¹² According to internalism, the feeling of pleasure can be characterized without reference to any external attitude; while according to the more plausible versions of externalism, pleasure is the feeling of having some experience to which some attitude external to it is being taken, such as a desire to sustain that experience. So Sidgwick's objection to the lack of correlation between strength of attitude and feeling of pleasantness is an objection to any such version of externalism. For, though he focuses on desire, the objection can be restated so as to apply to any attitude, including, for example, approving or welcoming.

Having rejected the volitional view, Sidgwick then goes on to consider a simple feeling-tone view according to which 'there is a measurable quality of feeling expressed by the word "pleasure", which is independent of its relation to volition, and strictly undefinable from its simplicity...like the quality of feeling expressed by "sweet", of which also we are conscious in varying degrees of intensity' (127).

Like many other writers, Sidgwick swiftly dispatches the feeling-tone view with what we might call the *heterogeneity objection*. When he considers pleasure in the broad sense—'to include the most refined and subtle intellectual and emotional gratifications, no less than the coarser and more definite sensual enjoyments'—he can find no single phenomenological property common to all pleasurable experiences. We shall shortly come to Sidgwick's own account of pleasure. But for the moment I submit that the heterogeneity objection moves in an unjustified way from the obviously true claim that the kinds of experience that are enjoyed are very various to the claim that the feeling of enjoyment is not some phenomenologically internal property common to all such experiences. On the face of it, it might appear that there is such a property. We can ask of any enjoyable experience whether it felt enjoyable, or whether it was more or less enjoyable than some other experience, and it is quite plausible to suggest that answers to these questions do not involve our attending to any attitude or other state external to the enjoyment or lack of it in question. This, combined with the difficulties encountered by versions of externalism (including Sidgwick's, as we shall see), provides the basis for a strong case for the feeling-tone view.

As I have previously argued (2006: 109–10), the feeling-tone theorist would be best advised not to appeal to an analogy between enjoyment and

12. I myself have preferred not to characterize the positions I discuss using this terminology, since in the literature externalism is often described as the view that pleasure has no distinctive phenomenology (e.g. Sumner 1996: 90).

a special sensation such as sweetness. Enjoyment is better understood as a determinable, with particular kinds of enjoyment—that in eating, reading, thinking, and so on—as determinates. So the heterogeneity argument is structurally similar to the argument that, because the visual experience of, say, seeing red is quite different from that of seeing blue, they have nothing phenomenologically in common, and so must be conceptualized by reference to some external attitude or other state. Pleasure is a feeling-tone, but bodily enjoyment in, say, a massage is as different from a mental enjoyment, such as reading poetry, as is seeing red from seeing blue.

2.3 The Apprehension View

Having rejected the feeling-tone view, Sidgwick continues:

the only common quality that I can find in the feelings so designated seems to be that relation to desire and volition expressed by the general term 'desirable', in the sense previously explained. I propose therefore to define Pleasure—when we are considering its 'strict value' for purposes of quantitative comparison—as a feeling which, when experienced by intelligent beings, is at least implicitly apprehended as desirable or—in cases of comparison—preferable. (127)

What is meant by 'implicit' apprehension? The contrast is most probably with cases in which the subject self-consciously recognizes and makes explicit to herself, or to others, that the feeling in question is desirable. To insist on explicit apprehension would be patently absurd, since such self-conscious recognition is so rare. But note also Sidgwick's restriction to the experience of intelligent beings. This suggests that it may well not be his view that what *makes* an experience pleasant is its actually being apprehended (perhaps only implicitly) as desirable. That would have the obviously implausible implication that non-intelligent beings could not experience pleasure.¹³ But of course his avoiding that implication leaves him with the question of just what it is that makes an experience pleasant.

13. At 129, Sidgwick removes the reference to intelligent beings entirely, defining pleasure as 'desirable feeling, apprehended as desirable by the sentient individual at the time of feeling it'. See also 131: 'Let, then, pleasure be defined as feeling which the sentient individual at the time of feeling it implicitly or explicitly apprehends to be desirable'. De Lazari-Radek and Singer (2014: 243) refer to the passage at 131 as Sidgwick's 'fullest and most precise definition of pleasure'.

To which notion of ‘desirable’ is Sidgwick here referring? He is presumably referring back to 1.9 (see Schneewind 1977: 317), where his final view is that the desirable is what ‘I should practically desire if my desires were in harmony with reason’ (112)—in other words, what I *ought* to desire (see M3 126 n). On this view, then, pleasures are such that, if experienced by an intelligent being, they would—perhaps only implicitly—be apprehended as good.

But questions arise here about *why* that being would take the view in question were there not some important phenomenological property of her experience that might justify her desiring it, a property independent of the experience’s actually being desired by her or its being such that it would be desired by her. Further, Sidgwick seems to be assuming that any informed and intelligent being must accept that pleasure is good. On the face of it, we can fairly easily imagine some intelligent and informed ascetic, who believes that all that matters in life is self-realization, contemplating some enjoyable experience and finding it worthless. Sidgwick’s response at 129 to a similar difficulty is to suggest that the ascetic must accept that the judgement that a feeling is desirable is implied in its recognition as pleasure, but can go on to claim that philosophy shows such a judgement to be mistaken. But it is not clear why the ascetic cannot deny this outright, seeking a purely non-evaluative, phenomenological account of pleasure, on the value of which a judgement can subsequently be made.

2.4 The Desirable Consciousness View

At yet other times, Sidgwick drops not only the reference to intelligent beings, but also that to the very idea of apprehension, speaking merely of ‘desirable consciousness’ (397; 398; 402; 404). Such passages suggest that apprehension may have been introduced as part of an attempt by Sidgwick to capture the epistemological point he wishes to make about the privileged access individuals have to the quality of their own feelings considered merely as feelings, which itself is what he appeals to in excluding non-hedonistic evaluations of experience (see e.g. 128). At 398, Sidgwick defines the ultimate good as desirable consciousness and continues with a reference back to 2.2:

According to the view taken in a previous chapter, in affirming Ultimate Good to be Happiness or Pleasure, we imply (1) that nothing is desirable except desirable feelings, and (2) that the desirability of each feeling is only

directly cognisable by the sentient individual at the time of feeling it, and that therefore this particular judgment of the sentient individual must be taken as final on the question how far each element of feeling has the quality of Ultimate Good.

Here there is no reference to any special sense of ‘desirable’, so I presume that we should take it in its usual sense—that is, as equivalent to ‘good’. Once again, however, we—like the ascetic—may wish to object that the project of explaining the nature of pleasure is quite different from that of evaluating it. Further, the definition again raises the issue of exactly what it is about these feelings that *makes* them desirable. Sidgwick may be able to sidestep the objection that we, like the ascetic, may make judgements about feelings based on non-hedonic properties (127–8). But when it comes to the hedonistic judgement, Sidgwick appears to commit himself to the idea of pleasantness understood independently of its being merely desirable:

if I in thought distinguish any feeling from all its conditions and concomitants—and also from all its effects on the subsequent feelings of the same individual or of others—and contemplate it merely as the transient feeling of a single subject; it seems to me impossible to find in it any other preferable quality than that which we call its pleasantness, the degree of which is only cognisable directly by the sentient individual. (128; see 94; 398–9; 400–1)

In the absence of any reference to an external state such as a desire, attitude, or cognition, it is tempting to read Sidgwick as here reverting to what amounts to a feeling-tone position, according to which what pleasurable experiences have in common, and what makes them valuable, is their having the special quality of feeling pleasant.¹⁴ I suggest, then, that Sidgwick at heart accepts the feeling-tone view of pleasure, but is misled by the heterogeneity argument into developing various forms of externalist account which are open to objection but in the end disappear from his theory.

3 Hedonism

3.1 Psychological Hedonism

Sidgwick begins his chapter on ‘Pleasure and Desire’ (1.4) with a highly perceptive survey of various kinds of moral motivation, noting the point

14. Note Sidgwick’s readiness in FTDA 104 to speak of pleasurable and painfulness as matters of ‘feeling-tone’.

often ignored in debates on this issue that an individual's motivation to act rightly may depend on her philosophical position (see CD I.3). So a rational egoist may be driven by self-love, a Kantian by reverence for authority, and a virtue ethicist by aspiration or admiration. But he then notes that one account of human motivation is often thought to bring to an end any debate about the principles on which action ought to be regulated: *psychological hedonism*, according to which 'volition is always determined by pleasures or pains actual or prospective'.

Sidgwick points out, however, that this view is quite consistent with the claim that a person's pleasures and pains may depend on her conception of right and wrong. That conception may be non-egoistic, and so there is no necessary connection between psychological hedonism and rational egoistic hedonism (the view that the right end for me is my own greatest happiness). Nevertheless, it is true that one who accepts psychological hedonism in the Benthamite form, according to which each person always aims at her *greatest* happiness overall, cannot conceive of reason's requiring her to aim at anything inconsistent with her own greatest happiness. Further, a Benthamite psychological hedonist cannot claim even that she *ought* to pursue her greatest happiness, or that it is a *dictate of reason* that she do so, since on Sidgwick's view any such requirement must be one from which the agent can deviate.

Sidgwick might have pointed out that, on his own view about rightness, such a psychological hedonist could still claim such action to be right or reasonable. But he goes on to point out that the psychological hedonists of his own day would admit the possibility of a person's knowingly acting against what she believes to be the overall balance of her pleasure over pain. For them, egoistic hedonism expressed in the language of *ought* is an option. Indeed, it seems a plausible option. Because of this, and because another common view is that virtuous action in its highest sense cannot be performed solely for the agent's pleasure, psychological hedonism must be carefully examined.

The first question Sidgwick grapples with (42–51) is whether the psychological hedonist is right to claim that all desires are for pleasure or to avoid pain. (We have to assume that Sidgwick is speaking here of non-instrumental desires. A psychological hedonist would have no difficulty explaining my desire for an ice cream, on the assumption that I see it as a means to pleasure.) He notes the oddity of Mill's suggestion that 'desiring a thing, and finding it pleasant, are, in the strictness of language, two modes of

naming the same psychological fact' (1998: 4.10), given that he also asks us to decide whether we desire only pleasure through introspection. Sidgwick's analysis is that, like Hobbes and others, Mill is confusing the idea of a person acting 'as she pleases' with the idea of her acting for the sake of pleasure. Mill's substantive view is that we desire only pleasure and the absence of pain.

Against Mill, Sidgwick sets the view of Butler,¹⁵ according to whom we could not pursue pleasure at all unless we had desires for objects other than pleasure, since pleasure consists in the satisfaction of such 'disinterested' desires (again, we have to assume that the non-hedonic desires in question are non-instrumental). Sidgwick notes Butler's exaggeration, drawing attention to desires for pleasures such as those of sight, hearing, or smell, which do not seem to depend on previous desires. Nevertheless, he claims: 'as a matter of fact, it appears to me that throughout the whole scale of my impulses, sensual, emotional, and intellectual alike, I can distinguish desires of which the object is something other than my own pleasure' (45).

The first example Sidgwick gives is hunger, which he plausibly interprets as a non-instrumental desire to eat. Eating is often pleasurable, of course, but hunger is not a desire for that pleasure. And there is often pleasure in the anticipation of eating, but it is the object not of appetite itself but of a 'secondary' desire. Nor should we allow that hunger can be made sense of by psychological hedonism as an aversion to pain. Aversion to pain is a desire to move from one's current state into some other construed merely negatively as not containing the present pain; whereas the desire for food is for moving into a state construed positively (as involving food). Again, there is often a secondary desire in play—to avoid the pain of ungratified desire; but that is not the same as the primary desire itself, as can be seen from the fact that the secondary desire can lead to the attempted suppression of the primary.

This leads Sidgwick into a general discussion of whether all desire is to some degree painful.¹⁶ Using hunger again as his example, he plausibly

15. 44 n. 2 attributes the same view to Hutcheson and Hume.

16. A long note appended to 1.4 (54–6) develops the argument against Bain's view that desire is painful, drawing a helpful distinction between the 'unrestfulness' of both desire and pain, and the 'uneasiness' of the latter (56), and warning us against the view that any strenuous activity is painful. This note is partly based on FTDA, a paper written primarily as a response to H. R. Marshall; several arguments from it do not find their way into the *Methods*.

suggests that it is not, pointing out that in many cases an unthwarted primary desire is a component in a highly pleasurable state of consciousness. Consider in particular the ‘pleasures of pursuit’ as against the ‘pleasures of attainment’. In the case of games, for example, people often enter into them primarily because they desire the pleasures of pursuit; the desire to win emerges only as they become involved in the game itself. And the same phenomenon is found when someone enters into useful social or scientific work for the sake of occupying herself.

Unlike the primary and secondary desires involved in hunger, those involved in pursuit and attainment can be inconsistent with one another (48–9). This is the so-called ‘paradox of hedonism’. In the case of ‘active enjoyments’ generally, we cannot obtain them fully if we aim at them consciously. Consider, for example, the pleasures of business, study, or creative art: ‘in order to get them, one must forget them’. Sidgwick is here advertent to the phenomenon now known as ‘flow’ (Csikszentmihalyi 1991). What about benevolence and sympathy (49–50)? We do of course feel sympathetic pains and pleasures, and helping others is often pleasurable. But the motive to beneficent action produced by sympathy is often much stronger than any desire for sympathetic pleasure or the alleviation of sympathetic pain. Sidgwick is honest about his own *schadenfreude* and self-regard:

Often indeed we cannot but feel that a tale of actual suffering arouses in us an excitement on the whole more pleasurable than painful, like the excitement of witnessing a tragedy; and yet at the same time stirs in us an impulse to relieve it, even when the process of relieving is painful and laborious and involves various sacrifices of our own pleasures. Again, we may often free ourselves from sympathetic pain most easily by merely turning our thoughts from the external suffering that causes it: and we sometimes feel an egoistic impulse to do this, which we can then distinguish clearly from the properly sympathetic impulse prompting us to relieve the original suffering.

He concludes the section by pointing out the analogy between benevolence and other desires requiring ‘disinterestedness’. To enjoy benevolent action requires a prior desire to help others for their own sake and not for our own. But here there need be no paradox of hedonism. Love often involves a strong desire for reciprocation, and the consciousness of one’s own self and its pleasures and pains is often increased by the intensity of affection. Nevertheless, there can be self-suppression, and it can even result in ‘acts of fantastic and extravagant self-sacrifice’. But we should remember that, if the pleasure of virtue can be obtained only if it is not consciously

sought, it will be just one example of many such pleasures: 'base and trivial external ends may excite desires of this kind, as well as the sublime and ideal' (51).

The third section of 1.4 interestingly foreshadows Sidgwick's later discussions of the 'dualism of practical reason'. He begins by cautioning against excessive stress on the incompatibility of 'self-regarding' and 'extra-regarding' impulses, suggesting that one's greatest happiness is attained through 'a sort of alternating rhythm' of each. Conscious desire is usually extra-regarding, but the agent, especially when desire is strong, is often keenly aware of the pleasures in play. So the conflicts between self-love and extra-regarding desire (the most prominent of which constitutes the conflict between virtue and the agent's pleasure: 52), when they do arise, should not be surprising, since both elements are constantly present. So if we are continually acting from impulses aimed at objects other than our own happiness, it is not surprising if we sometimes, as in cases of weakness of will, give in to those impulses when they conflict with our overall happiness. Or consider those cases in which people sacrifice their own happiness for posthumous fame, or some ideal end such as 'Truth'. Self-sacrifice is not only possible, but unsurprising.

The final, fourth section of the chapter considers two objections to Sidgwick's position. The first is the suggestion that overall pleasure is the *unconscious* aim of all human action. Sidgwick responds that consciousness alone appears to provide the only evidence of what we aim at, and that any unconscious 'aim' can only be some organic state conducive to self-preservation or the preservation of 'the race'. Neither of these is a sufficient response, since it is easy to conceive of behavioural experiments which might be used to test the hypothesis, and there seems no reason why unconscious desire might not be directed at pleasure if it secures, for example, motivational resources that might advance the survival prospects of the agent or her genes. But the onus here is on the objector, not on Sidgwick, since Sidgwick is surely right to say that we should rely on the evidence of introspection until counter-evidence is provided.

According to the second objection, our original impulses were aimed at pleasure and the avoidance of pain, and any other impulses are derived from these by 'association of ideas'. Again, Sidgwick cannot find the evidence for the claim. Children, he claims, show the dualism of motivation he has identified in adults, and indeed are often more driven by extra-regarding motivation, since they lack the reflective capacity necessary for consciously

aiming at pleasure. And even were all our original appetites for pleasure, this would have little bearing on our current situation, in which people desire many things other than pleasure.

3.2 Against Synonymic Hedonism

It was thought by some in the nineteenth century that the word ‘good’ is a synonym for ‘pleasant’. Sidgwick takes this view, which we might call *synonymic hedonism*, seriously (106–9), noting that—if we restrict ourselves to non-moral goodness—our judgements about goodness correlate closely with those about pleasantness. Indeed, this correlation is something he himself will use in his later argument for evaluative hedonism. But, he suggests, when we focus on attributions of goodness in the sphere closest to that of the assessment of conduct—the aesthetic sphere—the correlation we find does not support synonymic hedonism. For attributions of goodness here line up with *specific kinds* of pleasure—that of goodness in wine, for example, with the pleasure of tasting it, not with any hedonic effects of its being health-promoting. So the most we could expect in the case of the moral assessment of conduct would be a reference to some kind of ‘moral’ pleasure, analogous to the aesthetic pleasure that accompanies the attribution of beauty to some object. In fact, Sidgwick continues, even this will not be forthcoming. In the aesthetic sphere, we accept the judgements of aesthetic value of those with good taste, not those who claim, truly, that some object gives them pleasure. And the judgements of good taste imply some universal standard.

Sidgwick appears to be assuming that this standard must be non-hedonistic. In fact, the defender of synonymic hedonism might argue that the universal standard itself concerns aesthetic pleasure. So the pleasure experienced by non-connoisseurs is not genuinely aesthetic; that kind of pleasure can be experienced only by the person of good taste. But what Sidgwick goes on to say blocks this response. Connoisseurs of aesthetic objects can often continue to make sound aesthetic judgements even when they are jaded and no longer gaining pleasure from the objects they are assessing. Given that we have no reason to think that such connoisseurs are misunderstanding language, ‘aesthetically good’ cannot mean ‘aesthetically pleasant’. At this point, the synonymic hedonist would be forced to introduce the notion of a standard observer into the very meaning of ‘good’, claiming that ‘aesthetically good’ means something like ‘would be pleasant to an informed and

non-jaded observer'. Sidgwick does not discuss this view; but we can set against it his earlier implication that judgements about aesthetic goodness appear to concern non-hedonic properties, such as beauty, rather than pleasantness, whether actual or hypothetical. Indeed, he himself again refers to the idea of such a universal and non-hedonistic standard in his summary at 109 of the dense preceding paragraphs. In persons of 'good moral taste', he suggests, 'we can distinguish the intellectual apprehension of goodness—which involves the conception of an ideal objective standard—from the pleasurable emotion which commonly accompanies it; and may suppose the latter element of consciousness diminished almost indefinitely'.

The concluding paragraph of 1.9.2 (109) is historically significant, as well as puzzling. Sidgwick shifts attention from the adjectival to the substantive use of 'good', noting that the hedonist claim that pleasure is the good clearly implies—because it is not a tautology—that the meaning of 'pleasure' and 'the good' are different, even if they denote (i.e. refer to) the same thing. As we saw (CD 1.2–3), this passage represents one of the anticipations in Sidgwick of G. E. Moore's famous 'open question argument'.¹⁷ Sidgwick concludes the section with the terse statement that 'it does not seem that any fundamental difference of meaning is implied by the grammatical variation from adjective to substantive'. Given this, and the obvious power of the brief argument in this paragraph, one wonders why Sidgwick felt the need for the complicated discussion in the preceding three paragraphs. Hedonists will say—non-tautologically—that a life is good in so far as it is pleasant; and this simple point is enough to derail the synonymic hedonist understanding of 'good'. Indeed, by introducing the idea of non-hedonic sources of goodness, Sidgwick has stirred up trouble for his own later hedonistic arguments.

3.3 Welfare Hedonism

According to *rational egoism*, the only reason I have to act in any way depends on the extent to which that action furthers my own good, well-being, or welfare.¹⁸ But, to become practical, any such version of egoism must be

17. Moore 1903: 15–17. Note the reference to 1.3 of the *Methods* at 17 n. See CD, chapter 1 nn. 11, 20.

18. Sidgwick himself objects to the first two notions as empty at 91–2. But—as his reference forward to 1.9 and 3.13–14 suggests—he recognizes that particular conceptions of a person's 'good' can be provided so as to enable egoism, for example, to be distinguished from non-egoist views such as utilitarianism. In this passage, Sidgwick goes on to note two uses of 'happiness' that differ from his: those meaning a calm state of 'agreeable consciousness', distinct from specific pleasures, or a non-hedonistic conception of a person's good.

combined with an account of what that good consists in. According to what I shall call *welfare hedonism*, the only positive constituent of well-being is pleasure or enjoyment, and the only negative constituent is pain or suffering, a natural implication of the view being, of course, that the best life for any individual consists in that life with the greatest balance of pleasure over pain. Welfare hedonism combined with rational egoism will give us the view to which most of book 2 of the *Methods* is dedicated: *egoistic hedonism*. Welfare hedonism is consistent with the view that there are goods other than pleasure, and ‘bads’ other than pain, as long as these goods and bads are not seen as constituents of well-being. So I might believe that, though a person’s life can be improved only through an increase in the overall balance of pleasure over pain within that life, a *world* can be improved through, say, an increase in its beauty (beauty being understood as a good in itself, independently of its being appreciated by any sentient being), or a more equal distribution of well-being. According to *global hedonism*, there are no such non-hedonic goods or bads. Sidgwick accepted both welfare and global forms of hedonism.¹⁹ The only ultimate good is desirable consciousness (397). As does Sidgwick, I shall concentrate in particular on welfare hedonism, though it is worth noting that the arguments for and against welfare hedonism often carry across directly to the global form of the theory.

Sidgwick’s arguments for hedonism need to be understood in the context of his overall epistemological position, which will be discussed further in chapter 4. One major component of that position is his philosophical intuitionism, according to which certain propositions are ‘self-evident’ and a person who properly and reflectively grasps them can be justified in believing them on the basis of that grasp. I shall suggest that Sidgwick does seek to provide an intuitionist basis for his hedonism.²⁰ Like many hedonists, Sidgwick tends not to begin with positive arguments for the hedonistic position before moving on to consider alternative positions and objections. Rather, he outlines his arguments in response to these positions and objections.

We have already seen (CD 1.2) that, partly on hedonistic grounds (91), Sidgwick rejects the view that human well-being consists in self-realization,

19. I assume that passages elsewhere—for example, in Sidgwick 1998—which imply a belief in non-hedonic values are cases of Sidgwick’s employing the language of common-sense morality, which on his view is anyway founded on hedonistic utilitarianism.

20. At 98, Sidgwick states clearly that hedonism must be based on a ‘moral intuition’. See McTaggart 1906: 408; Frankena 1956: 542–4; Skelton 2008: 191–3.

understood as adherence to one's original nature. At 391–2, he argues against the view (of 'early moral reflection') that 'general good' consists in 'general virtue' (see Hurka 2001: 8–9; Shaver 2008). His argument begins by referring back to the claim defended in the earlier chapters of book 3 that 'virtue' (i.e. conformity to common-sense morality) is itself grounded on the promotion of the ultimate good on the whole of the agent, or the good of others. So to claim that the good consists in virtue is circular, since determining virtue itself depends on determination of the good.

This objection seems quite right. Such a view makes no sense without some independent content for the notion of well-being. Nor, Sidgwick goes on, does it help to add a notion of 'trained insight' (i.e. practical wisdom) to one's conception of virtue: the reference to good will remain (392–3). One might attempt to draw a distinction between virtue as a disposition and the acts required by common-sense morality, claiming that the good consists in such a disposition. Sidgwick's objection here (393–4) is that the disposition in question can be valuable only because of the acts or feelings it produces. This assertion is questionable: many will see moral value in the mere possession of, say, courage or kindness, independently of any actual manifestations of those dispositions. But Sidgwick could still appeal to his argument from circularity: exactly which acts or feelings it is good for the agent to have a disposition towards will itself depend on common-sense morality and hence what is good for individuals generally.

At 394–5, Sidgwick considers the claim that circularity can be avoided by identifying the ultimate good with *subjective* rightness (that is, the will to do the action believed right), understood independently of any knowledge of *objective* rightness and its basis in ultimate good.²¹ As Sidgwick points out, this of course avoids the circle he has described, since that requires acceptance of the view that rightness—that is, following common-sense morality—is itself grounded on promotion of the good. But it also seems absurd to claim that ultimate goodness depends solely on willing what is right, regardless of what one wills. Whether one's will is or is not tracking what is objectively right must matter. Further, common sense seems to hold that there are times when it is better not to be motivated by 'conscious obedience to . . . conscience'.

Having asserted that, as far as he can tell, *all* components of human excellence or perfection, not just virtue, are valuable merely for the 'good or

21. As Schneewind (1977: 313) notes, Sidgwick almost certainly has Kant in mind here.

desirable conscious life' in which they consist or which they will promote, Sidgwick faces the question whether well-being consists in desirable consciousness, with virtuous action as one component, and perhaps other elements also, including 'physical action, nutrition, and repose' (395–7). He first deals with the physical processes, plausibly claiming that, while they are of course instrumentally important, they cannot have any value in themselves. He concludes: 'In short, if a certain quality of human Life is that which is ultimately desirable, it must belong to human Life regarded on its psychical side, or, briefly, Consciousness.' This conclusion can be accepted, as long as Sidgwick is taken to be claiming merely that a physical process cannot, considered purely as a physical process, serve as a constituent of well-being, and not that well-being consists only in conscious states. Nothing that he has said here shows, for example, that those physical processes which constitute one's accomplishing something cannot, in so far as they are accomplishments, constitute well-being. Consider, for example, the movements of Mozart's hands over his manuscripts.

At 395–6, Sidgwick plausibly gives short shrift to 'zoological conceptions' of well-being, according to which it consists in tendencies towards self-preservation or the preservation of the subject's 'community or race': if all life were painful, there would be nothing but badness in its preservation. He then goes on to consider whether virtuous activity might be a component of well-being. Here Sidgwick merely states his view, without argument (Shaver: 2008: 220–1):

[S]o far as we judge virtuous activity to be part of Ultimate Good, it is, I conceive, because the consciousness attending it is judged to be in itself desirable for the virtuous agent; though at the same time this consideration does not adequately represent the importance of Virtue to human wellbeing, since we have to consider its value as a means as well as its value as an end. (397)

Sidgwick goes on to clarify the means/end distinction in this context by imagining a case in which a virtuous life is extremely painful, asserting that modern—as opposed to ancient—thinkers would not accept that such a life could be good overall for the virtuous agent, but might allow that it was the agent's duty to suffer that pain for others, or that periods of pain, though bad in themselves, might be instrumentally valuable for the agent (Sidgwick may have in mind, perhaps, the idea of a reward in the afterlife). As clarification of the means/end distinction, the argument here is largely unobjectionable. But Sidgwick provides no argument for the claim that modern thinkers will see periods of painful virtue as overall bad for the subject; and the example

certainly provides no evidence that virtue could not be *a* good for the agent, even if its value is outweighed by the disvalue of pain (see Irwin 2009: 459–60).

Sidgwick begins the following section (398–400) by asking whether we should identify desirable consciousness with pleasure. He goes on to refer back to his claim in 2.2²² that the ultimate good is happiness or pleasure, and to elucidate two of its implications: (1) nothing is desirable except desirable feelings; (2) the desirability of a feeling is directly cognizable only by the sentient individual at the time of feeling it, so that her judgement of desirability or pleasantness (see 398 n. 2) must be taken as final. He claims that no one will dispute the claim that feelings can be evaluated *as feelings* only in terms of their pleasantness.

Sidgwick nowhere offers a definition of ‘feeling’. At 404, however, he appears to be using the term as equivalent to ‘consciousness’. Clearly he does not accept this equivalence, however, since he allows in 398 that there are *other* elements of consciousness than feelings—cognitions and volitions. So we may conclude that a feeling is any element of consciousness which is *not* a cognition or a volition. Given the breadth of the notion, it seems unlikely that all will agree that pleasantness is the only property relevant to the evaluation of feeling. Consider, for example, the sense of awe experienced by a woman gazing at her newborn baby. She might well describe this as one of the most important moments of her life, and as a major contributor to her well-being. The awe, here, could be seen as *intentional*: it is awe *at* what she sees before her. But that does not make it a cognition, and—even though she might well enjoy the moment greatly—she might also refer to other good-making properties, such as profundity, or novelty.

When we consider the value of cognitions and volitions within consciousness, Sidgwick suggests, we should consider them independently of any feeling that may accompany them and of any relation they may have to their object (in the case of cognitions) or to an objective ideal (in the case of volitions). If we do that, he suggests, we shall see the value in consciousness of *mere* cognition or *mere* desire to be neutral. If we then allow in the relation to the object or an objective ideal, we might be tempted to think that, say, the cognition of truth or volition to realize freedom is non-hedonically valuable. Here Sidgwick again refers us back to 2.2.2 (125–30). He has in

22. He almost certainly has in mind 128.

mind the following important passage, in which he is responding to just the kind of complaint about potential non-hedonic good-making properties as that in my previous paragraph:

When . . . we judge of the preferable quality (as ‘elevation’ or ‘refinement’) of a state of consciousness as distinct from its pleasantness, we seem to appeal to some common standard which others can apply as well as the sentient individual. Hence I should conclude that when one kind of pleasure is judged to be qualitatively superior to another, although less pleasant, it is not really the feeling itself that is preferred, but something in the mental or physical conditions or relations under which it arises, regarded as cognisable objects of our common thought. For certainly if I in thought distinguish any feeling from all its conditions and concomitants—and also from all its effects on the subsequent feelings of the same individual or of others—and contemplate it merely as the transient feeling of a single subject; it seems to me impossible to find in it any other preferable quality than that which we call its pleasantness, the degree of which is only cognisable directly by the sentient individual. (128; see 113)

But it is not clear why any individual should not claim to be the final arbiter of whether she is experiencing a feeling of awe or profundity, without reference to any ‘common standard’.

On the basis of that earlier argument, Sidgwick concludes that, though we may appear to value the cognition of truth or a volition to realize freedom, what we are really valuing is their effects on future consciousness or ‘something in the objective relations of the conscious being’ (399). Sidgwick goes on to bolster the argument from 2.2.2 with further illustrations of what he means by ‘objective relations’. Someone’s preference for the state of apprehending truth over that of relying upon fictions is not a preference for merely a state of consciousness, but for a certain relation between the mind and reality. Someone who prefers freedom to slavery prefers this not because freedom is more pleasant, but because she is averse to a relation of slavery between her will and that of another. And if a person prefers conformity to virtue or the contemplation of beauty to more pleasant states of consciousness, this is because she is relying on the idea that her conception of virtue or beauty corresponds to an ideal that is ‘to some extent objective and valid for all minds’ (400).

The most problematic example here is that of the contemplation of beauty, which is analogous in various ways to my case of awe. Consider someone contemplating Michelangelo’s *Pietà*. There seems no reason why she should not claim that her consciousness at that time is non-hedonically valuable

by reference to intrinsic qualities of the experience itself, such as the appreciation of the serenity of the body of Christ. No metaphysical commitment to objective standards of beauty is required, just as, in the case of pleasure, Sidgwick assumes no commitment to any objectively correct principle concerning the value of pleasure.

But, as Sidgwick recognizes, a non-hedonist may anyway suggest that we *include* the ‘objective relations’ themselves in our account of well-being, in such a way that cognition of truth, perhaps, is valuable for the subject in a way that merely apparent cognition of truth is not. Sidgwick has already considered such so-called ‘ideal goods’ in 1.9, asserting that ‘it will...be generally held’ that they are valuable only in so far as they contribute to happiness or perfection (114). In 3.14.5 (400–7), he provides a two-part argument against idealism about welfare.

His first argument appeals to the intuitive judgement of the reader herself, describing the result of his own reflection as follows:

[T]o me at least it seems clear after reflection that these objective relations of the conscious subject, when distinguished from the consciousness accompanying and resulting from them, are not ultimately and intrinsically desirable; any more than material or other objects are, when considered apart from any relation to conscious existence. Admitting that we have actual experience of such preferences as have just been described, of which the ultimate object is something that is not merely consciousness: it still seems to me that when (to use Butler’s phrase) we ‘sit down in a cool hour’, we can only justify to ourselves the importance that we attach to any of these objects by considering its conduciveness, in one way or another, to the happiness of sentient beings. (400–1)

The power of this first argument, such as it is, depends entirely on the degree to which Sidgwick’s experience is mirrored in that of his readers. And, of course, many of them will disagree with him, as Sidgwick himself notes as he begins his discussion of his second argument (401; Skelton 2008: 205–6). As we shall see in the following chapter, Sidgwick himself provides a strong argument that, in such cases of disagreement, suspension of judgement is required. Further, Sidgwick misses the point, brought out clearly by Moore (1903: 27–3 and *passim*; see Seth 1896: 422; Bradley 1935: 91), that objective relations may play a role in constituting valuable elements of well-being even if, when considered in isolation, they are valueless. So if we consider, say, an individual’s knowledge independently of any conscious state with which it might be connected, it may well appear valueless. But

this is not to say that such knowledge, when combined with such states, constitutes a good which is not merely that of the states themselves.

Sidgwick's second argument consists in an appeal to 'a comprehensive comparison of the ordinary judgments of mankind' (400). Here we see, rather than Sidgwick's philosophical intuitionism, his Aristotelian commitment to the role of dialectic in ethics (see e.g. Aristotle 1894: 1145b2–7; also Irwin 2009: 489). Consider, for example, Aristotle's own discussion of happiness. In a central chapter of his *Nicomachean Ethics* (1894: 1.7), Aristotle offers his famous 'function' argument to the conclusion that the human good consists in virtuous rational activity. In the following chapter, he attempts to shore up that conclusion in part by showing how his conclusion is consistent with, indeed supported by, many of the 'ordinary judgments of mankind'.

Against the non-hedonists, Sidgwick notes that the ideal goods do produce pleasure in several ways, and that common sense approves of them roughly in proportion to the degree of such productiveness (401). He claims that this is obvious in the cases of beauty and freedom. As noted above, he is rather too sanguine about this. For a clear denial of Sidgwick's position on beauty, for example, consider Moore's case of the beautiful universe, which Moore believes has value, even if it is never seen by anyone (1903: 83–4). Sidgwick accepts that knowledge is harder for the hedonist to deal with, but notes first that common sense is especially impressed by knowledge that bears fruit. Even when it is the result of 'blue skies research', it may receive some approbation on the grounds that such apparently pure research often has unforeseen useful results, that it provides the enquirer with the innocent pleasures of satisfying curiosity, and that the intellectual disposition towards the acquisition of such knowledge is likely to be valuable as a whole. Further, just as common sense tends to set limits on the acquisition of knowledge according to its fruitfulness, so the same is true of the question of how far another important alleged ideal good, virtue, is to be pursued in independence from its valuable hedonic consequences (402; see also 406). As Shaver points out, this argument fails to show that virtue is not a good, as is true also of the arguments concerning knowledge (2008: 224; 226).

Sidgwick must and indeed does accept, however, that common sense is not, in the end, hedonistic. But he adduces four considerations which are intended to explain this and hence take the sting out of this aspect of common sense (402–6), and one can see his approach as analogous in various

ways to the ‘debunking’ strategy he employs against common-sense moral principles (see CD 3.4.2):

- (1) The word ‘pleasure’ tends to suggest the ‘coarser’ feelings, whereas the scope of welfare hedonism extends to all kinds of enjoyment (see 93; 120–1). Also, because certain pleasures often involve greater pain, or the loss of greater pleasures, we are reluctant to include them in our account of the ultimate good, especially as we often have moral or aesthetic concerns about them.
- (2) Many pleasures can be felt only on condition that we desire things other than those pleasures themselves. (Here Sidgwick refers back to 1.4 and 2.3.)
- (3) Common sense tends to be averse in particular to the ‘narrow and limited’ end of *egoistic* hedonism.²³ (Sidgwick’s thought here is that common sense fails properly to distinguish egoism from hedonism.)
- (4) Universal happiness is also likely to be better achieved if we restrict the degree to which we aim at it consciously. First, more limited ends are more achievable. Second, each person, if she is to be happy, needs ends, to be sought for their own sake, other than the happiness of others (and these may include virtue, truth, freedom, beauty, and so on, which of course may also have hedonically valuable consequences).

Sidgwick ends this key chapter of the *Methods* by raising the question whether there is any coherent alternative to evaluative hedonism. What we require is a standard for comparing not only ideal goods against one another, but ideal goods against happiness hedonistically construed (on the assumption that such happiness must be at least a component of any plausible conception of the ultimate good). Unsurprisingly, Sidgwick tells us that he knows of no systematic alternative to universalistic hedonism (and of course we can understand him also to be ruling out any egoistic alternative to egoistic hedonism). He is unwilling to allow as sufficiently systematic any pluralistic account of the good which relies not on any single reductive standard for comparison, but on informed judgement about the relative value of tokens of different types of good. But this view has much to be said for it, and—since it is widely accepted by ‘cultivated persons’—it is another matter on which suspension of judgement is called for by a central principle in Sidgwick’s own epistemology.

23. For further discussion of 403–5, see CD 7.1.1.

3.4 Empirical Hedonism

Merely believing that the good or well-being consists only in the balance of pleasure over pain is not enough to make one an evaluative hedonist in the fullest sense. For this view, as stated, leaves open the question what *makes* happiness, so understood, good. It is consistent, for example, with the view that the only good-making property is that of promoting self-realization, and that the only route to self-realization is through maximizing the balance of pleasure over pain. A fully hedonistic view will allow only one good-making property—pleasantness—and one bad-making—painfulness. It is clear that Sidgwick is a full hedonist in this sense, and indeed he sees the position as required to make coherent the very idea of maximizing happiness:

pleasures should be sought in proportion to their pleasantness; and therefore the less pleasant consciousness must not be preferred to the more pleasant, on the ground of any other qualities that it may possess. (121)

It is because he is a full hedonist that Sidgwick rejects J. S. Mill's famous suggestion (1998: 2.4–5) that hedonism can allow pleasure to be evaluated in terms of quality as well as quantity (94–5).²⁴ Indeed, Sidgwick goes so far as to claim that the purely quantitative position follows from the definition of pleasure itself: 'all pleasures are understood to be so called because they have a common property of pleasantness, and may therefore be compared in respect of this common property' (94).

Though he has a clear view on the quality/quantity distinction, Sidgwick is less clear on the distinction between egoistic hedonism as a theory of ultimate reasonableness, on the one hand, and as a set of practical principles, on the other:

We must . . . understand by an Egoist a man who when two or more courses of action are open to him, ascertains as accurately as he can the amounts of pleasure and pain that are likely to result from each, and chooses the one which he thinks will yield him the greatest surplus of pleasure over pain. (121)

This is again a result of Sidgwick's failure to properly distinguish principles and methods (see CD 1.4), though it is particularly surprising given his

24. I have argued elsewhere that Mill's is a possible version of hedonism (Crisp 1997: 31–5), but it does involve the somewhat peculiar view that properties such as nobility correlate with pleasantness and—despite apparently being valuable—do not themselves contribute to a person's well-being.

painstaking and insightful demonstration of how certain pleasures can arise only when not sought directly. Nevertheless, it is highly unlikely that the most plausible version of egoistic hedonism would be entirely ‘self-effacing’ (Parfit 1984: 23–4; see also M7 129–30). There will be times when subjects are required to make decisions on the basis of hedonistic assessments of their options, and Sidgwick devotes a good deal of space to the question of the viability of such assessments. This is primarily in his discussion of *empirical hedonism* in 2.2 and 2.3.

Sidgwick begins 2.1 by clarifying what one is committed to by the very idea of the ‘greatest happiness’: ‘we must assume the pleasures sought and the pains shunned to have determinate quantitative relations to each other; for otherwise they cannot be conceived as possible elements of a total which we are to seek to make as great as possible’ (123; see 94–5). Note that Sidgwick is not claiming that empirical hedonism requires a cardinal, as opposed to a merely ordinal, measure of value. Indeed, he goes on to note the possibility of ‘discontinuity’—that there are some pleasures of which the smallest amount would outweigh the greatest conceivable amount of some others—and the fact that it would only simplify hedonistic calculation (see CD 7.2.1).

There is a fascinating footnote to this passage, referring to a discussion of discontinuity by Edmund Gurney (123 n; see Gurney 1887: 1.4):

[I]n the case of pain, it has been deliberately maintained by a thoughtful and subtle writer, with a view to important practical conclusions, that ‘torture’ so extreme as to be ‘incommensurable with moderate pain’ is an actual fact of experience... This doctrine, however, does not correspond to my own experience; nor does it appear to me to be supported by the common sense of mankind:—at least I do not find, in the practical forethought of persons noted for caution, any recognition of the danger of agony such that, in order to avoid the smallest extra risk of it, the greatest conceivable amount of moderate pain should reasonably be incurred.

Consider the most appalling agony imaginable, and try to compare it with some very mild discomfort—a dull, almost unnoticeable twinge in one’s little finger, say. Many people are prepared to take seriously the thought that *no* amount of finger-twinge could be worse than, say, a year of the excruciating agony. What Sidgwick notes is that, on the standard view of rational choice, those who take this view should be prepared to accept an infinite period of finger-twinge (or perhaps even something worse) to avoid the

tinest risk of agony.²⁵ Neither Sidgwick nor Gurney discusses the question of discontinuities involving both pleasure and pain. It strikes me as a view worth serious consideration that there is some period of agony, considerably shorter than a typical human life, such that *no* amount of pleasure of whatever kind could counterbalance it in value. Consider, say, fifty years of exquisite torture. If that is right, given Sidgwick's point about probability, the rational course of action for many human beings would be to seek painless extinction as soon as possible, so as to avoid the risk of such torture. That conclusion seems absurd; and yet the underlying theory of value does not.²⁶

Sidgwick plausibly goes on to suggest that the relevant variables in quantitative comparison are the duration and the intensity of pleasure, that we should assume a 'hedonic zero' point from which the positive value of pleasure and the negative value of pain can be measured, and that we should deny the Epicurean paradox that pleasure is just the absence of pain (124–5).

The third chapter of book 2 is a brilliantly insightful, sustained, and honest examination of the practicality of empirical hedonism. In recent years, many of Sidgwick's points and doubts have been borne out by empirical research in psychology and related disciplines, and so-called 'positive psychology' still has much to learn from Sidgwick.²⁷ What then is the 'empirical-reflective method'? It is:

to represent beforehand the different series of feelings that our knowledge of physical and psychical causes leads us to expect from the different lines of conduct that lie open to us; judge, which series, as thus represented, appears on the whole preferable, taking all probabilities into account; and adopt the corresponding line of conduct. (131)

To make such comparisons more manageable, Sidgwick suggests, we might ignore manifestly imprudent conduct, as well as less probable and minor contingencies.

25. Gurney's original 1881 essay, cited in M3 119, was revised and reprinted in the collection to which Sidgwick refers in M7. In the revised version, he includes a spirited response to Sidgwick (Gurney 1887: 183–4), claiming, plausibly enough, that 'the effect on the mind of a very high probability of immunity may practically be the same as that of certainty'.

26. For further helpful discussion see Temkin 2011: ch. 5.

27. The Sidgwickian programme has been imaginatively developed by Kahneman (e.g. 1999; 2000; see Kahneman et al. 1999b).

Sidgwick states the obvious fact that people *do* make comparisons of pleasures on the basis of intensity, but notes that such comparisons may be said to be too unusual and rough to constitute a system, and to be potentially erroneous (140–1). Sidgwick denies the Platonic suggestion that the bodily pleasures are misunderstood as positively pleasurable, being merely states of relief from prior pain: here, the immediate judgement of consciousness is decisive (141). But there are nevertheless serious potential difficulties. Merely focusing on a pleasure may diminish it; and any estimate of its intensity will require comparing it with some other past state, which introduces the danger of misrepresentation by one's memory or, if one is relying on the views of others, mistaken testimony (141–4). Further, it is difficult to compare two pleasures experienced contemporaneously since, first, their causes may interfere with the pleasantness of each (so no general conclusion could be drawn about the intensity of each kind), and, second, they are hard to distinguish.

Sidgwick provides some personal illustrations of the general difficulties he has raised for empirical hedonism. He says that, in the case of a dinner, he often finds it impossible to say whether one dish or wine gives him more pleasure than some other. Or, in comparing bodily or mental activities, it is only when differences are large that ranking is possible. He cannot rank kinds of pleasure, such as those of labour against those of rest, especially when they are 'impure' and involve some degree of unpleasantness or pain. For example, it is hard to compare a state of quiet satisfaction with one of hopeful suspense. And it is yet harder to compare pure pleasures and pure pains.

There are further reasons to suspect error (144–6). An individual's judgements of the value of certain pleasures vary over time, partly because of the representation of the feeling in question. Emotional pains, for example, are more easily recalled than sensational ones: I can more easily remember the unpleasant anticipation of seasickness than the nausea itself, though the latter was considerably worse; and the excitement of certain hardships leads us to remember them as more pleasant than they really were. But variation is also partly the result of the individual's own state of mind at the time of the comparison. So we exaggerate the pleasures of appetite when we desire them, and belittle them when sated. Fear can distort our judgements of future pain. And the experience of one kind of pain can lead us to play down other kinds. In a situation of danger, we might put a high value on repose, forgetting how tedious it could become. Likewise, when absorbed in

a certain pleasure, we can underestimate others, which itself poses a problem for comparisons involving pleasure experienced at the time of the comparison. Indeed, any kind of focused excitement can cause bias. Further, we find it hard to imagine the intensity of some pleasure which at present we cannot experience, such as that of exercise when considered at the end of a weary day. Nor is the 'cool hour' always the answer, since such a state may not be ideal for evaluating pleasures requiring enthusiasm or excitement.

These difficulties lead to the conclusion that the assumption of commensurability made by the empirical hedonist is just that—an assumption (146–7). And even if there is commensurability, and we allow that each pleasure and pain has some definite degree of pleasantness or painfulness, the question remains whether we can measure such degrees with any accuracy. It may well be the case that we are *always* biased for or against certain pleasures or pains, the particular bias in question depending on our state at that time. We can learn about possible sources of error and try to make allowances for them, and also make judgements of the same pleasure or pain at different times and in different moods. But even then all we can hope for is a 'rough approximation to the supposed truth'.

In addition, we have to remember that *we* could have changed, so that even if our judgement about some past pleasure or pain is correct, we cannot assume that our current experience would be the same (147–50). We may be satiated by a certain kind of pleasure, for example; and hedonistic calculations that work for a youth may be inappropriate once she is older. Nor will relying on the experience of others protect us from error. Human beings differ greatly from one another, which undermines Plato's suggestion that philosophy is more pleasurable than sensuality because the philosopher prefers it. And the more our advisers have to rely on their own memories, the greater the chance of mistake. Further, there is a real danger that our understanding of the attitudes of others to certain pleasures and pains will colour our estimates of our own. It is also often unclear to us how far we could go in the direction of changing ourselves so as to become more open to certain pleasures.

Sidgwick concludes 2.3 with the claim that all these considerations taken together must weaken our confidence in the 'empirical-reflective' method of egoistic hedonism. It is not that we should reject it outright. Sidgwick admits that he continues to make practical comparisons between pleasures and pains. But, he suggests, it would be helpful if we could call also upon some other method.

3.5 Objective and Deductive Hedonisms

What alternative might there be to empirical hedonism? In 2.4, Sidgwick discusses *objective hedonism*, according to which the egoistic hedonist can avoid the difficulties of empirical-reflective comparisons through relying on common-sense estimates of objects often sought as sources of pleasure.

It is not clear why Sidgwick distinguishes objective from empirical hedonism, since he has already suggested the testimony of others as a potential source of information for the empirical hedonist (142; 147–50).²⁸ The difference is perhaps partly between reference to common sense as part of a general empirical strategy, and complete reliance on it; but also important is that objective hedonism focuses on pleasant *objects* rather than pleasant feelings. This might itself be thought a methodological advantage, and the objective hedonist might also hope for some kind of convergence through relying on the combined experience of humanity, in which the differences between individual points of view have to some extent cancelled one another out (151).

Much of 2.4 is taken up with a sustained critique of objective hedonism so understood. Some of Sidgwick's objections to objective hedonism, as we might expect, are similar to those he has already made to reliance on common sense as part of an empirical strategy. But others are new.²⁹

Common sense, Sidgwick suggests, provides us only with estimates of pleasantness and painfulness for a typical human being, and most people live their lives anyway within a narrow compass, 'labouring to avert starvation and severe bodily discomfort' (151–2). Common sense is liable to the same distortions as empirical hedonism, and in addition it fails to distinguish between objects of natural desire and sources of pleasure, which of course are not strictly correlative (152). Common sense is influenced by moral and aesthetic judgement (153). And should it be guided by people's stated preferences, or their actions (153)? On the one hand, people are weak-willed, and may fail to aim at what they sincerely believe will bring them the greatest pleasure overall; on the other, the non-virtuous may overvalue virtue through a conscious or unconscious desire to gain moral credit.

28. At M1 136, Sidgwick explicitly notes the overlap. He sows further confusion by speaking of 'deductions' at M1 151.

29. It is worth bearing these arguments against common sense in mind when assessing Sidgwick's position on common-sense morality; see CD 6.9.

Common sense lacks clarity and consistency, even if we restrict ourselves to a single culture at one time (153–8). There is some consensus on the following as sources of happiness: health, wealth, personal relationships, fame and position, power, interesting work, and amusement, understood to include the satisfaction of the love of knowledge and the arts. But common sense provides no clear ranking of these goods. Even more worryingly, not only is the consensus not complete, but people are all too willing to admit that their own judgements on happiness are mistaken. So the wealthy expend a great deal of time and money on satisfying bodily appetite, but will tell us that ‘hunger is the best sauce’, or even, because of the adaptation of preferences, that wealth is often of little or no benefit to its possessors. Similar attitudes are commonly expressed to social rank, reputation, power, and high society. What about ‘domestic affections’? The truth is that, when men can satisfy their sexual desires without the burdens of a family, so many remain unmarried that legislation may be required. Aesthetic contemplation and intellectual activity are agreed to be valuable, but only a small minority has really experienced them, and again there is no agreement on their ranking vis-à-vis other sources of happiness or indeed on the ranking of different cultural activities against one another. Even if common sense were more consistent, there would still be important dissenting voices: religious ascetics, pessimists, philosophers. Should we listen to the philosophers because their reflection is the most careful? Again, we face the problem that they may be biased towards intellectual pleasures.

It is true that the commonly agreed sources of happiness rarely conflict with one another, and that the widespread quasi-philosophical doubts about them may be a reaction to exaggeration of their positive qualities (158–60). So an increase in wealth, for example, *usually* will increase happiness, up to a point. But in the end, common sense hardly claims to provide more than ‘rather indefinite general rules, which no prudent man should neglect without giving himself a reason for doing so’ (160).

After a chapter criticizing the claim that happiness consists in the performance of duty (see CD 7.2.2), Sidgwick ends book 2 with a discussion of another possible alternative to empirical hedonism: the *deductive* method (see 121–2). If I believe that God has promised happiness as a reward for obedience to certain rules, for example, I do not need to calculate the different levels of pleasure and pain in each of two options, one of which involves a breach of these rules. Or if I hold that the greatest balance of pleasure over pain will always follow upon the healthiest activities, I have

not to calculate the overall balance of pleasure over pain in various options, but merely to work out which activity is the most healthy and then ‘deduce’ that this constitutes my best option. Of course, deductive hedonism may be based on empirical judgements, indeed on empirical judgements about pleasure and pain. But, unlike empirical hedonism, it does not require case-by-case hedonistic assessment of possible options.

2.6 is Sidgwick’s attempt to deal with the approach of late nineteenth-century psychology to the question of pleasure and pain. There is again some overlap with his discussions of empirical and objective hedonism, but much of the chapter consists in ad hominem responses. Sidgwick recognizes that deductive hedonism is likely to be based on inductive inference from judgements about particular cases, but the hope might be that these judgements would be easy and that scientific principles based on them would suffice for harder cases also. The problem, however, is that there is no satisfactory general theory of the causes of pleasure and pain (178). One difficulty is that one class of pleasures has physical causes, another psychical (178–9). In essence, Sidgwick’s complaints about specific psychological theories of his day are as follows. Hamilton’s Aristotelian account, modified by Stout, according to which pleasure is ‘the reflex of...unimpeded exertion’ fails to cover bodily pleasures, and pleasures—such as those in games—that involve overcoming obstacles (180–3). Spencer retains the idea of ‘exertion’, suggesting that pains are excessive actions of organs, while pleasures correlate with moderate organic activity.³⁰ But it is not clear that pains always result from *excessive* organic activity so much as activity of a certain *kind*, and anyway such a theory would be hard to apply (183–7). Stout puts forward the hypothesis that we see pleasurable organic processes as restorations of ‘equilibrium’. This may cover such cases as relief after pain, but again many pleasures are left out. And if we extend the notion of equilibrium so that it covers any satisfaction of desire, we will be committed to the implausible view already discussed that desire is always painful (187–9; see 45–7; note at 54–6).

Sidgwick at this point notes the particular problems aesthetic pleasures pose for all of the theories he has so far discussed (189–90). In the next section (190–2), he discusses the ‘biological’ view of pleasure and pain, again found in Spencer, according to which pain correlates to activity that damages the

30. For an argument that Spencer’s hedonism is closer to that of Sidgwick’s empirical version than Sidgwick realizes, see Weinstein 2000.

organism, and pleasure to beneficial activity. But Spencer himself admits that this can be at best a generalization and that many sources of pleasure and pain are cultural rather than biological in the required sense. Nor can we link pleasantness to ‘quantity of life’ (192–5). A more ‘intense’ life may be more painful, and if we interpret quantity in terms of ‘harmony’, the theory will become hopelessly vague, partly because *concentration* on one source of pleasure can often increase it. Nor can we conclude that spontaneity will be a more reliable guide than reflection to overall pleasure. In the end, indeed, we have to accept that there is no ‘high priori road’ to the goal of maximizing individual happiness. Empirical hedonism is all we have.

4 Objections to Empirical Hedonism

4.1 Practical Objections

At 132, Sidgwick suggests that examining various objections to the practicability of (empirical) hedonism is a helpful way to clarify the theory itself.

The first we might call the *impossibility of conception* objection—Green’s claim that ‘pleasure as feeling, in distinction from its conditions that are not feelings, cannot be conceived’ (132–3).³¹ As Sidgwick notes, this contradicts not only common sense and the views of empirical psychologists, but an assumption which underlies many of Green’s arguments elsewhere. If Green means merely that pleasure cannot exist apart from aspects of experience that are not feelings, his view is undeniable; but this is irrelevant as to whether pleasure can be assessed independently of those conditions.

A second objection is the *non-additivity* argument, again found in Green (see also Bradley 1935: 85, discussed in Schneewind 1977: 394–6), that pleasures cannot be added since they occur in series and not at the same time (see 407 n). But if Green were right, then we could not add periods of time, since they occur in series. At times, Green appears to mean only that we cannot experience happiness all at once, which is of course something a hedonist will not deny. This does not prevent the notion of maximum happiness serving as a criterion of conduct.

31. Sidgwick responds to Green on hedonism in GSM lecs 7–8.

It may be that what Green means is that consciousness of the transiency of pleasure undermines its pleasantness or causes later pain (134–7). To this, Sidgwick responds robustly that it does not appear to him that way, as long as he believes himself to have a good prospect of similar pleasures in the future. It is true that we are likely to value permanence in the sources of pleasure; but that is quite consistent with hedonism. It is true also that consciously aiming at such sources may be self-defeating in hedonistic terms, but this is just the ‘paradox of hedonism’, which can be resolved in practice.³²

Sidgwick notes that it is sometimes said that there are certain higher impulses, such as the love of virtue, with which self-love is especially incompatible (137–8). But, he suggests, this difficulty is not recognized in standard moral philosophy or theology. It may be that in some very unusual cases the extreme pleasure which comes to someone who entirely abnegates her self would be obtainable only through her entirely giving up any kind of self-concern. But because any such case would be so rare, it cannot provide the basis for a serious objection to empirical hedonism.

A related criticism we might call the *reflection* objection (138–40). On this view, reflectively examining our pleasures itself diminishes them. We need not see consciousness in a zero-sum way, as if increasing attention decreases capacity for feeling. But there is a real danger here. It arises, however, particularly in the case of reflection upon intense pleasures while they are being experienced. The solution is just to avoid this kind of reflection, and to examine such pleasures from a later vantage point.

Sidgwick then moves to consider the difficulties in comparing pleasures (see CD 3.4).

4.2 Theoretical Objections

Empirical hedonism rests, of course, on evaluative (welfare or global) hedonism. The most common class of *theoretical* objection to any form of evaluative hedonism is that there are non-hedonic goods or values. For example, it is commonly said that there are certain pleasures, such as those of a sadist, which are so far from being valuable as to be *evil* (Harsanyi 1982: 56), or that hedonism is the *philosophy of swine*, in that it cannot make room for ‘higher’ aesthetic or moral values (Mill 1998: 2.3). One of the most

32. Sidgwick here refers back to 1.4.

influential objections to hedonism in modern times revolves around the example of the *experience machine*. The most famous statement of the objection is by Robert Nozick:

The Experience Machine. Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain... Would you plug in? *What else can matter to us, other than how our lives feel from the inside?* (Nozick 1974: 42–3; see 1989: 104–8)

The objection has most force in cases in which authenticity seems significant. Imagine that I am drinking a cold beer. It is not implausible for a hedonist to claim that the value of the enjoyment in this activity would be the same if I were in fact plugged into an experience machine. But now imagine that I am writing a novel, and enjoying it. According to the hedonist, the level of well-being of someone on an experience machine which ‘copies’ my veridical experiences and replays them to that person is, in this respect, equal to mine. This seems highly counter-intuitive. Don’t creativity, genuineness, authenticity, knowledge, truth, accomplishment, and many other goods matter *independently of pleasure*?

The experience machine objection, then, essentially amounts again to the claim that there are non-hedonic or ideal goods. So, although Sidgwick never confronts any of these objections explicitly, what he says about ideal goods provides us with the opportunity to construct a response on his behalf. As we have already seen, he draws attention to the fact that ideal goods are productive of pleasure, and claims that common sense approves such goods roughly in proportion to that productiveness. Further, he might appeal to his later ‘debunking’ account of common-sense morality, suggesting that the moral principles concerning evil, nobility, authenticity, and so on, on which such objections rest, are themselves grounded on the hedonic value which their adoption promises.

Nevertheless, as we saw above, Sidgwick’s epistemology appears to commit him, in the case of all objections relying on the notion of ideal goods, to suspension of judgement, since many reflective thinkers accept that the ideal goods illuminated by the objection are genuine, and there is no plausible account available of how such thinkers could be mistaken. Sidgwick may indeed *claim* that they are mistaken, believing that items valuable because productive of pleasure are valuable in themselves and misled by the fact

that pursuing these items *as if* good in themselves may—because of the paradox of hedonism—be the most effective way to advance the balance of pleasure over pain. But this claim would be a mere hypothesis which, though it may be open to empirical verification, has not been so verified or even properly tested.

4

Intuitionism

1 Intuition and Intuitions

As we have seen (CD 1.2), Sidgwick was not interested in providing a substantial metaphysical backdrop for his epistemological intuitionism. According to Sidgwick, intuition is a doxastic faculty (4–5; 28; 275), nothing more, or less, than a capacity for forming beliefs of a certain kind, with the possibility thereby of acquiring knowledge. These beliefs, because of their apparent source in intuition, may be called ‘intuitions’, and Sidgwick feels himself entitled to assume their *existence*, though not their *truth* (210–12):

I wish therefore to say expressly, that by calling any affirmation as to the rightness or wrongness of actions ‘intuitive’, I do not mean to prejudge the question as to its ultimate validity, when philosophically considered: I only mean that its truth is apparently known immediately, and not as the result of reasoning. I admit the possibility that any such ‘intuition’ may turn out to have an element of error, which subsequent reflection and comparison may enable us to correct; just as many apparent perceptions through the organ of vision are found to be partially illusory and misleading: indeed the sequel will show that I hold this to be to an important extent the case with moral intuitions commonly so called. (211; see 34 n. 2)

To this extent, then, Sidgwick’s intuitionism is cognitivist. Intuitions for him are not to be understood as, say, certain sentiments or attitudes, but as non-inferential (or ‘immediate’ (98)) rational beliefs which present themselves to the subject as true (and hence as potentially false)—that is, as apparently self-evident (see e.g. 338).¹ Each person has to make up her own mind about whether she has any such beliefs, and at this point Sidgwick is characteristically

1. This appearance can vanish on reflection: see e.g. 339; 343; 383. If some proposition appears to him clearly true in itself, and it survives the tests for ‘certainty’ outlined at 338–42, Sidgwick will describe it as ‘really self-evident’ (see e.g. 496).

cautious, noting that, in his experience, people seem likely to identify as moral intuitions non-cognitive states such as mere impulses to, or unspecific preferences for, action, or non-intuitive cognitive states, such as conclusions from rapid or semi-conscious processes of reasoning or familiar and unreflectively held opinions. The fact that Sidgwick denies the status of intuition to the latter shows that mere non-inferentiality is insufficient for a belief to be an intuition: it must also present itself as self-evident to the subject. If as soon as the subject holds her customary belief up to the light it seems not true to her, then that belief was and is not a moral intuition.

Some find talk of a faculty of intuition mysterious (Mackie 1977: 38–9), and may seek to undermine the idea through some developmental or evolutionary account (212–14).² Sidgwick refuses to allow that mere knowledge of the causal history of certain apparent cognitions provides any ground for distrusting them. All of our cognitive faculties have such a history, and hence so do all of our beliefs (including the beliefs of those who criticize intuitions). And the subject matter of ethical judgements is such that they cannot be inconsistent with any merely descriptive psychological or physical claims. Nor, even if our capacity for moral cognition had been present from the start, would that give it any special claim to correctness. None of this is to deny, of course, that some more specific causal account of the origin of certain beliefs might not increase the probability of their being mistaken. But we should remember that Sidgwick's metaphysical quietism does make his position less vulnerable to objections based on such accounts than certain more metaphysically substantive positions.

In the nineteenth century, intuitionists were often seen in opposition to so-called *inductivists*. Having presented a broad conception of intuition as 'immediate [i.e. non-inferential] judgement as to what ought to be done or aimed at', Sidgwick is careful to situate his intuitionism in the context of contemporary debate (97–8), noting quite correctly that the parties in the debate were commonly talking at cross purposes, since each was claiming to know different things. The inductivists claimed inductive knowledge of the pleasantness of certain actions, whereas intuitionists focused on the rightness (or wrongness) of those actions. Rational or normative hedonism itself cannot be known inductively. It must be either grasped intuitively, or inferred from other premises, at least one of which must include a normative intuition.

2. Sidgwick often speaks of the 'moral faculty', and at 34 he equates 'the faculty of moral cognition' with 'Reason'. That this is the source of intuitions is clear from e.g. 462.

What we see here is a standard argument for foundationalism in epistemology. Someone inclined towards a non-foundationalist approach, such as some form of coherentism, may well accept Sidgwick's negative argument against inductivism (that induction alone cannot ground an ethical theory), but deny that foundationalism is the most plausible alternative. Rather, it might be claimed, for example, that hedonism provides the most consistent and coherent fit with other beliefs we have about goodness, rightness, or the world in general. As we shall see (CD 4.2.3), Sidgwick's own relationship with coherentism is a complex one (cf. Hooker 2000a: 348).

Sidgwick goes on to clarify further the relationship between induction and his conception of intuitionism (98–9), pointing out something often ignored in late nineteenth-century ethics—that hedonism and inductivism are quite independent of one another.

Sidgwick's broad notion of intuition allows that intuitions can concern both rightness and goodness (103). Goodness here includes not only 'moral goodness' or virtue (105–6), but well-being. When Sidgwick is arguing for a hedonistic theory of well-being against positions which attribute value to certain 'objective relations' between a conscious subject and the world, it is the 'intuitive judgment' of the reader to which he appeals (400–1; see 406–7). Further, these objects of intuition may be seen as properties of acts or states of affairs (such as, say, the wrongness of this particular breaking of a promise), or as normative or evaluative propositions (such as 'it is wrong to break promises' or 'knowledge is good'). This distinction would not have concerned Sidgwick, since he would almost certainly have been prepared to see the first kind of intuition as equivalent to the grasp (or apparent grasp) of a proposition to the effect that this breaking of a promise is wrong. But the distinction between action and principle in general does underlie that he draws between perceptual or aesthetic intuitionism on the one hand, and dogmatic intuitionism on the other, as we shall now see.

2 Intuitionisms

2.1 Perceptual Intuitionism

In his chapter 'Ethical Judgments', Sidgwick recognizes that it is commonly held that 'the moral faculty' is concerned with individual actions and he admits that the faculty in question might be thought closer to a form

of 'moral sense' than one of rational intuition (33–4; see 228–9; 392). Just as my visual sense tells me that this is a table, so, it might be said, my moral sense tells me that this action is wrong; whereas it would seem less plausible to claim that my moral sense tells me that actions of this kind are wrong (or of course that my visual sense tells me that objects of this kind are tables). Sidgwick objects to this terminology that the term 'sense' suggests a capacity for feelings which can vary between two individuals without either being in error.

Sidgwick's thought here is perhaps as follows. If I put my hand in a bucket of water, my sense of touch may tell me it is warm; if you put your hand in the same bucket, your sense may tell you it is cold. But we might rationally agree to differ on the matter, concluding that the water feels 'warm to me' and 'cold to you'. We could not, according to Sidgwick, think of ethics in the same way, since ethical judgements involve 'the conviction, implicit or explicit, that the conduct approved is "really" right' (or of course wrong, if disapproved) (27).

But even such cases of apparent relativity might be disputed. It could plausibly be claimed that something is warm if it is such as to appear warm to a standard observer in normal conditions (Hardin 1983). In other words, in the bucket case, one of us is wrong, and perhaps explicably (I have been working in a freezer for five minutes). And often, of course, we are immediately happy to understand the deliverance of sense perception as veridical or not (the perception of my desk, say, is veridical, while that of the bent straw in my glass of lemonade is non-veridical). It may be that what really motivated Sidgwick to avoid the term 'moral sense' was the danger of its being mistakenly linked with some form of moral sentimentalism, as opposed to rationalism.

Later, in 1.8 ('Intuitionism'), Sidgwick uses the term 'conscience' for the faculty of moral judgement as applied to individual actions, noting that if in each case a person should follow their conscience so understood, there would be no need for general rules or 'scientific ethics' (99–100). Sidgwick calls this view *perceptual intuitionism* (102), describing it, in its most extreme form, as 'ultra-intuitional';³ calling it a 'method', indeed, is stretching the term a little, since the 'procedure' it involves is nothing more than a single

3. It is not clear what a less extreme form of perceptual intuitionism might be, since dogmatic intuitionism is that form of intuitionism which makes room for commonly accepted moral rules (100). Perhaps Sidgwick has in mind a hybrid form of intuitionism, which is partly perceptual in that it permits some judgements to be based purely on conscience alone.

judgement.⁴ He goes on to point out that few will accept such a position, since most people find their own particular intuitions open to doubt, non-comprehensive, inconsistent over time, and indeed often in conflict with those of others.⁵

In the ‘second’ ‘Intuitionism’ chapter (3.1), Sidgwick suggests two further problems for an ethics advocating a case-by-case assessment by conscience (214). Concrete cases are complex in a way that abstract principles are not, and often involve personal interests or sympathies. In these passages, Sidgwick expresses a clear preference for an ethics based on disengaged reflection in the ‘cool hour’, in such a way that the analogies he later seeks to draw between ethics and mathematics should come as no surprise. Against his position, of course, may be ranged all those anti-cognitivist arguments to the Pascalian conclusion that the heart has its reasons, of which reason knows nothing, as well as the point that many of the objections Sidgwick makes to particular judgements apply also to moral rules as commonly experienced: they too often seem open to doubt, insufficient always to provide guidance (partly indeed because of their very abstractness), and both intra- and interpersonally inconsistent. Sidgwick’s hope must be that he can show how his own conception of moral rules avoids these problems.⁶ As we shall see, it is interpersonal disagreement in particular which will prove the main stumbling block to his project.

2.2 Dogmatic Intuitionism

The focus of perceptual intuitionism is particular actions, and for that reason it might have been better called ‘particularist intuitionism’ or ‘action-focused intuitionism’. For Sidgwick distinguishes the next ‘phase’ of intuitionism through reference not to some further moral faculty, but to the object of the faculty in question—viz. moral rules. There is no reason why I should not claim to ‘see’ that promise-breaking is wrong in itself, just as I ‘see’ that this instance of promise-breaking is wrong (see de Lazari-Radek and Singer 2014: 69).

4. On so-called *aesthetic intuitionism*, see 228–9; 392–3. Aesthetic intuitionism is, in effect, a version of perceptual intuitionism, in which insight, in the light of a prior indeterminate account, is that into which actions realize virtue. As we shall see, Sidgwick appears largely to ignore such non-extreme versions of perceptual intuitionism (see CD 6.9).

5. As Phillips notes (2011: 93), we are reminded here of the conditions for certainty Sidgwick later outlines in connection with allegedly self-evident intuitions (338–42; see CD 4.3.1).

6. Sidgwick sees ethics as inquiry into the ‘right rules of conduct’ (4), and is prepared to describe his own view in terms of ‘rules’ (see e.g. P6 xviii–xix).

Further complications arise when we consider Sidgwick's initial, narrower conceptions of intuitionism (96–7).⁷ The first relies on the philosophical usage of his time, restricting the scope of intuition to rightness, and to the rightness of actions *in themselves*, independent of their consequences (see 7–8; CD 1.5). We might call it *non-consequentialist intuitionism*. Sidgwick is careful to point out that no morality has ever ignored consequences entirely, and that distinguishing an act from its consequences is often difficult. But such distinctions are part of common-sense morality, which therefore provides the material for a non-consequentialist conception of intuitionism. Sidgwick offers an example (see Butler 1865: sect. 79). I may intentionally tell the truth about some fact to a jury, knowing that they will, on the basis of what I say, form a mistaken view about the guilt of the accused. But their being misled is a mere consequence of my act of truth-telling, and does not prevent its counting as an instance of veracity.

Sidgwick then goes on to distinguish yet another version of intuitionism, which we might call *consequentialist non-hedonistic intuitionism*. According to non-consequentialist intuitionism, the rightness of acts is to be assessed as a property of those actions in themselves, independently of their consequences. Consequentialist non-hedonistic intuitionism concerns not the rightness of acts in themselves, but the goodness of their consequences. Its judgements, however, are also made on the basis of judging those consequences *immediately* and 'in themselves', independently of the pleasures and pains they involve, and it is primarily this feature which Sidgwick has in mind here.⁸

It is important not to allow Sidgwick's subcategories of intuitionism broadly construed to direct one's attention away from the core claim of intuitionism, which is epistemological: that we have the capacity to make 'immediate'—that is, non-inferential—judgements about rightness and goodness, which do not, as is the case with empirical beliefs, depend on evidence. If one loses sight of this, it will be hard to understand why he groups under the heading of 'intuitionism' so many different theories—in particular, dogmatic and philosophical intuitionism.

Dogmatic intuitionism emerges as a response to the alleged weaknesses within perceptual intuitionism (100–1). Sidgwick draws an analogy here

7. Note that both broad and narrower forms could be focused on either individual acts or on rules.

8. Sidgwick says that this view is intuitionist in a 'wider' sense than non-consequentialist intuitionism, perhaps because it concerns consequences of acts rather than just acts themselves; but we should note that in another sense it could be seen as narrower, since it does not assess the goodness of acts in themselves.

between moral and jural reasoning (see 2–4; CD 1.4). Just as an act will be pronounced lawful or unlawful by being considered in the light of the rules of positive law, so it may be seen as right or wrong by being brought under one or other moral rule. Mere commitment to such moral rules, however, is not on its own enough to make one an intuitionist, since one could accept them on the authority of others. What matters is their being seen as true (i.e. correct) and binding by the agent herself. According to the dogmatic intuitionist, general moral rules are implicit in common-sense moral thought, and the task of the philosopher is to elucidate and systematize them as far as possible. The importance of this can be seen again through reference to the legal analogy (228). If someone is to be under an obligation, whether legal or moral, he should be in a position to know exactly what that obligation is when he is making decisions. An indefinite law is a bad law, an indefinite moral principle a bad principle.

Dogmatic intuitionism is not, then, entirely unreflective; indeed, it is the result of reflection on the *dogmata*—the beliefs—of common sense. But Sidgwick no doubt felt the name appropriate partly because he believed the view to be *insufficiently* reflective, and hence unable to provide a coherent underpinning for common-sense morality itself, an underpinning which would enable the agent to know exactly which obligations she was under in each situation in which she found herself. What is particularly objectionable to Sidgwick about dogmatic intuitionism, then, is not its starting from common-sense morality, but its readiness to end there also without having removed the indefiniteness which is, he thought, as inappropriate in ethics as in a legal system.

2.3 Philosophical Intuitionism

Philosophical intuitionism emerges out of dissatisfaction with the morality of common sense (101–2), and it is hard not to see a sideswipe at Whewell in Sidgwick's implication that those satisfied with that morality lack a 'philosophic mind'.⁹ Those with such a mind, Sidgwick claims, will question not the general authority of common sense, but its lack of system. *Even if* common-sense morality can be made internally consistent, comprehensive, and such as to provide answers to all practical questions,

9. Grote is also in his sights. At GU 173, Sidgwick claims that 'to a philosophical mind it is only a "systematic" manner of thought . . . that can approve itself as a "right manner"' (see Schneewind 1977: 191).

still the resulting code seems an accidental aggregate of precepts, which stands in need of some rational synthesis. In short, without being disposed to deny that conduct commonly judged to be right is so, we may yet require some deeper explanation *why* it is so. (102)

A defender of dogmatic intuitionism might object that her theory does indeed provide this deeper explanation. Promise-breaking, for example, can be seen to be wrong in itself, or wrong because unjust; helping others is good in itself, or good because it is benevolent; and so on. For Sidgwick, however, such a theory is still unacceptably unsystematic. He is seeking 'one or more principles more absolutely and undeniably true and evident, from which the current rules might be deduced, either just as they are commonly received or with slight modifications and rectifications'.

On the face of it, then, Sidgwick appears to find two deep problems with dogmatic intuitionism. The first—which will take him most of book 3 to demonstrate—is that the rules of common-sense morality cannot in fact be made sufficiently precise, consistent, comprehensive, or decisive. The second is that even if they *could* be, they would still lack a background principle (or some—presumably small—set of principles) which would explain and justify them, thus showing them to be not merely 'accidental'.

Given the huge amount of attention given to the first problem, it is odd how little Sidgwick says about the second. The effect of this is that the weight of the argument against dogmatic intuitionism rests largely on the criticisms made of it in book 3. As we shall see, a dogmatic intuitionist may respond to many of these criticisms through introducing some suitably developed conception of practical judgement. But even with that conception in place, it is still open to Sidgwick to argue that dogmatic intuitionism is unsystematic. Disappointingly, all we have in the *Methods* at this point is assertion. If Sidgwick could have supplied a powerful argument for rational synthesis, he might, as I have suggested (CD Preface), have omitted from the *Methods* much of the detailed discussion of common-sense morality in book 3, restricting himself to demonstrating a clear lack of rational synthesis in one or two cases and how even a reflective form of dogmatic intuitionism would fail to provide such synthesis.

Having concluded that common-sense morality, even in its most developed form, will be unable to provide clarity and certainty, Sidgwick reintroduces philosophical intuitionism as follows:

[W]e conceive it as the aim of a philosopher, as such, to do somewhat more than define and formulate the common moral opinions of mankind. His

function is to tell men what they ought to think, rather than what they do think: he is expected to transcend Common Sense in his premises, and is allowed a certain divergence from Common Sense in his conclusions. It is true that the limits of this deviation are firmly, though indefinitely, fixed: the truth of a philosopher's premises will always be tested by the acceptability of his conclusions: if in any important point he be found in flagrant conflict with common opinion, his method is likely to be declared invalid. Still, though he is expected to establish and concatenate at least the main part of the commonly accepted moral rules, he is not necessarily bound to take them as the basis on which his own system is constructed. Rather, we should expect that the history of Moral Philosophy—so far at least as those whom we may call orthodox thinkers are concerned—would be a history of attempts to enunciate, in full breadth and clearness, those primary intuitions of Reason, by the scientific application of which the common moral thought of mankind may be at once systematised and corrected. (373–4; see 102; 214–15)

This significant passage illuminates further Sidgwick's attitude to dogmatic intuitionism. As he sees it, the view assumes from the start that common-sense morality is broadly justified. So, in a sense, dogmatic intuitionists are telling people 'what they ought to think': they ought to accept common-sense morality, as specified by the relevant dogmatic intuitionist theory. But they provide no supporting principle or principles for common-sense morality because they never attempt reflectively to 'transcend' that morality.

The charge as expressed here is odd, since Sidgwick himself seems prepared to assume that common-sense morality has sufficient initial authority for it to provide limits to the acceptability of any moral theory that 'transcends' it.¹⁰ Further, on the face of it, there seems no reason why a dogmatic intuitionist should not claim to have started with as open a mind as Sidgwick, and then to have found that the theoretical principles developed out of common-sense morality are both plausible in themselves as 'primary intuitions of reason' and able to provide a standpoint from which to systematize and correct ordinary moral thought (see e.g. Aristotle 1894: 1145b2–7;

10. At M1 472, Sidgwick speaks of 'our confidence in the substantial veracity of Common Sense'. See PM 31; Shaver 1999: 70. Skelton (2010a: 516–17) suggests that when Sidgwick claims that the method of the philosopher which conflicts with common opinion 'is likely to be declared invalid', he is merely making a prediction about what will happen as a 'word of caution' to philosophers. But above Sidgwick has said that, according to 'our' conception of a philosopher, any philosopher is permitted only a '*certain* divergence' from common sense (my italics). It also seems clear that, on this conception, the philosopher is 'expected' (that is, 'required', unless we also take the 'bound' that follows predictively) to provide a foundation for the main components of common-sense morality.

Ross 1939: 1–2; Phillips 2011: ch. 4; see Crisp 2014: 242).¹¹ That is, she may transcend common sense in the sense of impartially reflecting upon it, and yet find within common sense the kind of grounding principle or principles Sidgwick is seeking.¹²

The passage also raises the question of Sidgwick's own epistemology. Given his pessimism about the prospects of common-sense morality as a source of fundamental ethical intuitions, his awareness of its historical contingency and the variation in moral views across time and space (e.g. 360; 379), and the fact that his most basic normative questions can be asked without use of the conceptual scheme of common-sense morality (e.g. 'what do I have most reason to do?' rather than 'what ought I to do morally?', 'what obligations do I have?', or 'what is my duty?'), it is surprising to find Sidgwick allowing any significant epistemic role to common-sense morality.¹³ Why does he not advocate distancing oneself as far as possible from positive morality—that is, transcending it entirely—and seeking direct insight into reasons for action?¹⁴

Sidgwick's statements about the aim of a philosopher are somewhat indirect, but it is highly unlikely that they are not meant to apply to him. He himself

11. Note also that dogmatic intuitionists might well end up a long way from common sense by providing idiosyncratic interpretations of common-sense principles. Indeed, both Aristotle and Ross could be said to be in this category.
12. Note that Sidgwick himself is examining the methods of ethics he finds *within* common-sense morality, and in outlining the principles he finds self-evident makes frequent reference to common sense (379–84).
13. These reasons for doubting common-sense morality suggest that it does not represent to *Sidgwick* what he 'thinks he *knows*... about what he ought to do' (Schneewind 1977: 262; my italics). At 215, Sidgwick says that there would be 'apparent agreement at least among moral persons of our own age and civilization' on the 'validity' of the rules of common-sense morality. But believing that some rule is justified is not to believe that one's own belief in its justification is both epistemically justified and true and, as such (Gettier problems aside), equivalent to knowledge. Schneewind claims also that Sidgwick begins with the assumption that some 'complex set of moral rules is required for day-to-day rational guidance of action' (1977: 284), which might give common-sense morality some kind of default epistemic status. But I cannot find evidence for this view in Sidgwick's text, other than his argument in favour of *some* rule-based view against perceptual intuitionism. Nevertheless, we should remember Sidgwick's positive description of Socrates's 'general faith in the common sense of mankind' (OHE 29; see Schultz 2004: 60; 69). Further, Sidgwick clearly thought that common-sense morality *seems* to provide some fairly secure principles, such as those governing promising, and this certainly provides him with a reason for examining at least parts of it in search of genuinely self-evident principles. But, in the *Methods* at least, he might perhaps have started and finished with promising.
14. For the view that he did, and hence compares favourably with Rawls, see Singer, P. 1974. In my view, several of the criticisms Singer makes of Rawlsian reflective equilibrium apply to Sidgwick's position on common-sense morality. For further examples of Sidgwick's attaching epistemic weight to common sense, see e.g. the suggestion that he is not 'disposed to deny that conduct commonly judged to be right is so' (102); and the claim that the view of the good as desirable consciousness is 'in harmony with Common Sense' (396).

will test the plausibility of a philosopher's theory¹⁵ by checking that it will 'establish and concatenate at least the main part of the commonly accepted moral rules' (see Hooker 2000a: 348). But by doing so, he plays into the hands of the dogmatic intuitionist, while weakening his own positive case for the importance of self-evident and foundational intuitions in ethical theory. I suggest that what we see here is an unwillingness by Sidgwick to entirely surrender the methodology of Aristotelian dialectic,¹⁶ in which one's own philosophical conclusions must be tested against what is commonly thought.¹⁷ If he had done so, his theory would certainly have been more radical; but it would also have been epistemically purer and hence more systematic.¹⁸ Since Sidgwick nowhere explains exactly how much epistemic weight to give to common-sense morality—that is, how far a philosopher's conclusions may diverge from that morality—he is leaving it not only open for a dogmatic intuitionist to use the many cases in which utilitarian conclusions are violently at odds with common-sense morality as an internal argument against Sidgwick's own arguments for utilitarianism, but also unclear how we are to decide between two theories, one of which may have more self-evidence on its side but is less consistent than the other with common-sense morality.¹⁹

3 Self-evidence

3.1 Conditions for Certainty

Sidgwick is seeking normative principles that are genuinely self-evident—that is, true and evident when considered independently of other propositions (see VOB 124; EEFP 29). They need not be in any sense *immediately obvious*—that is, easy to grasp. This conception of self-evidence follows from

15. I am assuming that this is what Sidgwick means by 'method' in the passage quoted: a philosopher's premises and conclusions.
16. Aristotle 1958: 100b21–3. At PC 24, Sidgwick expresses a preference for the 'moral thought of the reflective few to that of the unreflective many'. The views of the reflective few would presumably be taken into account when testing one's beliefs against the non-dissensus condition (see CD 4.3.1), so there would be no independent justificatory role for common-sense morality (even when applying the second 'reflective' condition: see CD 4.3.1).
17. See Schneewind 1963: 140; 145; Schultz 1975: 350; Sverdlik 1985: 540–5; Brink 1994: 182–4; Schultz 2004: 9; 174–204; Irwin 2009: 467. For an interpretation of Sidgwick as a pure foundationalist intuitionist, see Skelton 2010a. For a defence of an epistemology involving foundationalist and coherentist elements, see e.g. Haack 1993.
18. See Sverdlik 1985: 548.
19. It is worth noting how rarely Sidgwick discusses cases in which utilitarianism and common-sense morality are radically at odds.

his understanding of moral judgement itself, according to which we make non-inferential judgements about the rightness and wrongness of certain acts. Non-inferentiality is not sufficient for self-evidence, however. If one is uncertain about some non-inferential judgement, it could not be said to be even apparently self-evident. To say that some claim is self-evident is to say, in part, that it seems extremely plausible. Nor are subjective certainty and non-inferentiality enough for *genuine* self-evidence. Some false propositions may appear to be self-evident, and this raises the question of how we might go about sifting these out.

Any apparently self-evident proposition, if it is to achieve the ‘highest degree of certainty’, has to meet, fully, four conditions (338–42);²⁰ and, we can safely assume, to the extent that any such proposition meets any of these conditions, it is more certain (and hence ‘genuinely’ self-evident, though not *necessarily* for that reason true).²¹ These conditions are as follows:

(I) *Clarity*. ‘The terms of the proposition must be clear and precise.’

Note that Sidgwick appears not to require that the proposition itself be either clear or precise. Presumably, if the terms of some proposition are precise, then the proposition in which they are embedded is likely to be so, if the process of precisification extends to the relations between those embedded terms. But there is no reason to think that clarifying and precisifying the terms of some proposition will make that proposition as a whole clearer. Indeed, it may make the proposition considerably more obscure. We should therefore emend Sidgwick’s condition so that it covers the proposition as well as its terms.

(II) *Reflection*. ‘The self-evidence of the proposition must be ascertained by careful reflection.’²²

20. As Phillips (2011: 61) points out, Sidgwick takes these conditions to be at least in part empirically based on the experience of avoiding past error. He need not be understood as postulating independent a priori principles of ‘rationality’, and would perhaps have been ready to see the conditions as structurally analogous to the rules of common-sense morality (see Crisp 2015b: 151).

21. Phillips (2011: 81) contrasts an extreme foundationalist view of the conditions, according to which meeting them adds no further warrant or justification to the belief in question, with moderate foundationalism, which allows passing the conditions some justificatory role. As I understand the conditions, they do provide further warrant for the belief, in so far as meeting them decreases reasons for doubt. See Shaver 1999: 64–7.

22. A puzzle might be raised here (see Hurka 2014: 131). If the reflection is to ascertain self-evidence, what is there left for the other conditions to do? Sidgwick must be read as claiming that the *apparent* self-evidence of the proposition must be *tested* by reflection. He will not allow that anyone could ever be *entirely certain* of a genuinely self-evident proposition (211; see CD 4.1). Hurka suggests that Sidgwick’s aim is to establish the apparent self-evidence of the relevant proposition, but this seems unlikely, since the proposition being tested is already seen as apparently self-evident. The test is for the highest degree of certainty.

An intuition is not a gut reaction or instinct, or some commonly held opinion, but a belief which to ‘careful observation’ presents itself as a dictate of reason. Much of the reflection in question will concern the origin of the intuition: Is it a mere product of my upbringing? Am I adopting it unthinkingly on the basis of societal authority? Does it rest on my subjective likes and dislikes? And of course much will concern the content of the proposition itself. Sidgwick is aware of the special importance of such reflection in ethics in particular. He notes that strong sentiments often transform themselves into apparent intuitions, and that we accept many moral principles largely on the basis of reliance on external authority. Through habit, we feel an impulse to conform to those principles without further justification, so it will be essential to test whether that principle stands up to reflection as a ‘clear intuition of rightness’.

Although Sidgwick does not say so explicitly, it is tempting to believe that he was thinking also of reflection on the implications of the apparently self-evident proposition.²³ So reflection would include Sidgwick’s dialectical testing against common-sense morality of apparently self-evident principles. Ultimately, I am seeking to vindicate my beliefs, to reassure myself that I believe P because it is indeed the case that P. In general, we can see here Sidgwick’s attempt to avoid Bentham’s and Mill’s criticism of intuitions as mere prejudices (see Schneewind 1977: 134; 179; Crisp 2014: 238–9).

(III) *Consistency*. ‘The propositions accepted as self-evident must be mutually consistent.’

In the 3.II analysis, it is unclear whether Sidgwick is speaking of logical or practical consistency. The first would rule out as self-evident, for example, the two logically contradictory propositions, ‘One should always be spontaneous—that is, non-reflective’, and ‘One should always be reflective’, if held in that form by a single person. Practical inconsistency would rule out not only such sets of propositions, but also, for instance, ‘One should always be kind’, and ‘One should always be just’, if a case can be imagined in which the demands of justice conflict with those of kindness. The conclusion of the *Methods*, in which Sidgwick discusses attempts to make egoism and utilitarianism practically consistent, suggests that he had the practicality test in

23. See the reference to examining ‘the consequences of adopting’ practical principles at 13–14; also APE 167.

mind, as does the fact that he claims to find such conflicts in many ethical writings, recognized by their authors.²⁴

(IV) *Non-dissensus*.²⁵

Since it is implied in the very notion of Truth that it is essentially the same for all minds, the denial by another of a proposition that I have affirmed has a tendency to impair my confidence in its validity . . . And it will be easily seen that the absence of such disagreement must remain an indispensable negative condition of the certainty of our beliefs. For if I find any of my judgments, intuitive or inferential, in direct conflict with a judgment of some other mind, there must be error somewhere: and if I have no more reason to suspect error in the other mind than in my own, reflective comparison between the two judgments necessarily reduces me temporarily to a state of neutrality. And though the total result in my mind is not exactly suspense of judgment, but an alternation and conflict between positive affirmation by one act of thought and the neutrality that is the result of another, it is obviously something very different from scientific certitude.

This important condition is, unlike the first three, stated indirectly.²⁶ In the context of 3.II.2 as a whole, it appears to be the claim that, to be of the highest certainty attainable, some proposition I accept must not be denied by any person whose opinion I have no more reason to think mistaken than my own.²⁷ But Sidgwick in fact goes on to spell out the *degree* of ‘certainty’ I am entitled to have in my own opinion, as opposed to that of the other person—and it is none at all. It is true that his expression here is apparently descriptive rather than prescriptive (‘has a tendency to impair my confidence’ rather than ‘should impair . . .’; ‘reduces me . . . to a state of neutrality’ rather than ‘should reduce . . .’). But I take it that Sidgwick is here intending to describe his own experience as the *reasonable* response to disagreement; otherwise, why would he mention it?

The final sentence above, however, I do take to be merely descriptive. Sidgwick surely cannot be recommending some kind of epistemic schizophrenia in which the contradictory affirmations are somehow kept apart. In fact, I suggest, what should emerge, on a proper understanding of

24. See also 360, in which Sidgwick speaks of the ‘collision’ between the different principles of common-sense morality, and the difficulties of determining their boundaries of application.

25. At APE 168, Sidgwick calls this ‘Social or Oecumenical Verification’, a third ‘verification’ in addition to the ‘Intuitive’ (corresponding to I and II above) and ‘Discursive’ (corresponding to II and III).

26. In earlier work, Sidgwick had put weight on the idea of general agreement rather than on that of non-disagreement. See e.g. M1 320–1; EEFP 31–2; CD 4.3.3.

27. Sidgwick is clearly thinking of epistemic peers. The fourth condition does not—*pace* Phillips (2011: 79)—require consistency with ‘ordinary moral opinions’.

the non-dissensus condition, is indeed suspension of judgement, alongside an affirmation of a mere appearance.²⁸ If two people who believe themselves to be in roughly the same epistemic circumstances are confronted by what one of them takes to be a redwing, the other a song thrush, they should suspend judgement on which of the two the bird is. But the bird's appearance to each as one or the other may not change (see Crisp 2006: 93; 2007a: 33; 2011a: 151).

It might be suggested that level of certainty should correlate with degree of consensus. Sometimes, the level of disagreement will itself provide a reason to suspect error (imagine that in my example all of the many other birdwatchers present claim to see a redwing). Often, however, it will not. Imagine that 51 per cent suggest redwing, 49 per cent song thrush. We do not have to get into the details of Bayesian epistemology to see that no one on either side could plausibly claim any significantly greater degree of certainty for their identification.

Sidgwick did not fully confront the implications of the non-dissensus condition for his own normative ethics. He does in 384–6 seek to show that his own allegedly self-evident axioms are in agreement with the views of equally reflective philosophers, and this could be taken as his looking for potential disagreement to test his views rather than agreement to support them (see CD 4.3.3).²⁹ But he fails properly to reflect upon the fact that there will be reflective thinkers who disagree with each of his alleged axioms, however weakly they are interpreted. Such disagreement, indeed, is implicit within his 'dualism of practical reason' itself: egoists will disagree with the principle of benevolence, while utilitarians will rebut the principle of prudence.

It remains a mystery why Sidgwick appears not to have recognized the sceptical implications of his non-dissensus condition. One possible explanation is that he believed that one of its implications was a form of scepticism which would have paralysed his philosophy. But this is not the case: he could have said everything he wanted to say, but on the understanding that

28. See Sextus Empiricus 1994: sects 15; 31; 151; 196; etc.; also Striker 2004: 16. It is worth noting that since the non-dissensus condition itself is controversial, its advocates may have to suspend judgement on it. But, as I explain in the main text, this does not prevent their continuing to advocate it. For an argument that proponents of morality can, unlike egoists, avoid the implications of the fourth condition through appeal to the value of moral understanding as opposed to mere knowledge, see Hills 2010: pt 3. As Shaver points out (2011: 659), it is not clear that utilitarians can employ this argument, since they put no special value on understanding.

29. Though note his summary of the fourth condition in the 'Contents', which remained unchanged from M1: 'We require of an Axiom that it should be . . . (4) supported by an adequate "consensus of experts"' (xxxiv).

what we were getting were Sidgwick's own reports from the fronts of reflection about how things appeared to him. Still, it must be admitted that the non-dissensus condition does 'open a door to universal scepticism' (see 509), and Sidgwick clearly found unwelcome the thought of that door's opening to him.

3.2 Sham-axioms

Before setting out his own fundamental ethical axioms, Sidgwick warns us against certain 'sham-axioms', which strike us as self-evident only because they are largely tautological (374–9). He begins with the following account of three virtues:

Wisdom: It is right to act rationally.

Temperance: It is right that the lower parts of our nature should be governed by the higher. (See 343–4; 369–70; 395.)

Justice: We ought to give every person his own.

This sounds informative. But, Sidgwick suggests, once we learn that acting rationally is doing what we see to be right, that the higher part of our nature is reason, and that a person's 'own' is that which it is right that she should have, we can see that the account is tautological:

Wisdom: It is right to do what we see to be right.

Temperance: It is rational to act rationally, and not give in to impulses to act irrationally.

Justice: It is right to give every person that which it is right that she should have.

But Sidgwick's elucidations of the original views are not the only ones available. Consider the following:

Wisdom: It is morally right to do what you see you have strongest reason to do.

Temperance: It is morally right that desire should be governed by reason.

Justice: We ought, morally, to give every person that which he is entitled to possess.

The correctness of each of these claims is an open question. Against the first, it might be urged that in certain cases it would be wrong to do what you see you have strongest reason to do, if, for example, doing that will involve your furthering your interest at the expense of that of others. Against the second,

an anti-rationalist may urge the rightness of following one's passions.³⁰ And against the third, someone who puts great weight on self-reliance may argue that it is up to each person to secure their own entitlements for themselves, without being in a position to call on the assistance of others.

Sidgwick does not attribute to any particular writer the views of the virtues he here criticizes. But he goes on to object to Plato and Aristotle in particular that, though their views may not be tautological, they do little more than explain the ethical problem to be resolved rather than provide any solution to it. Both agree that knowledge of virtue is a central aspect of ethics. But Plato cannot tell us more of each virtue than that it comprises (1) the knowledge of what is good in certain circumstances and (2) a harmony between the elements of our appetitive nature, such that any impulse emerging from it is in accordance with that knowledge. But we are not given that knowledge itself.

Sidgwick probably has the *Republic* primarily in his sights here, but we should remember other dialogues such as the *Protagoras*, one reading of which is as a recommendation of a version of hedonism similar in several ways to that which Sidgwick himself outlines in book 2. In the *Republic*, Socrates explains why ethical knowledge is beyond us: it requires knowledge of the Forms. But he also explains in considerable detail the education required for the acquisition of such knowledge. What we see at work here, perhaps, is Sidgwick's view that ethics should be freed as far as possible from metaphysics. Plato would of course have disagreed, and it is hardly a criticism of him that he fails to provide a non-metaphysical ethics.

Sidgwick's reading of Aristotle is also somewhat uncharitable:

Nor, again, does Aristotle bring us much nearer to such knowledge by telling us that the Good in conduct is to be found somewhere between different kinds of Bad. This at best only indicates the *whereabouts* of Virtue: it does not give us a method for finding it. (376; see Barnes 1976: 23–6; Irwin 2007: 204–6)

As Sidgwick knew, Aristotle himself raises the same potential objection to his doctrine of the mean:

But having grasped only this, someone would be none the wiser; for example, you would not know what sort of treatments to use on your body if someone

30. In his critical discussion of temperance at 343–5, Sidgwick claims that this anti-rationalist view, though not tautological, nevertheless lacks self-evidence. In CD 6.1, I provide an alternative Aristotelian account of temperance or self-control which seems neither tautological nor as implausible (in itself, or as an understanding of temperance) as a blanket recommendation of rational reflection.

were to say that you should employ those that medicine requires. (1894: 1138b29–32)

Aristotle's response is to explain the role of 'practical wisdom' or *phronēsis* in moral agency. Sidgwick would doubtless object to the idea that practical wisdom consists partly in a perceptual capacity to see what is called for in certain situations, and not merely in the application of fully articulated and explicit principles. But since, according to Aristotle, practical wisdom will enable us to see what virtue requires, occasion by occasion, he cannot be charged with a failure to provide a 'method': it is a version of perceptual or aesthetic intuitionism. He also tells us about how virtue and practical wisdom are acquired, through the habitual performance of virtuous actions (1894: 2.1–4). He cannot provide an ethical decision-procedure independent of practical wisdom; but that is because he believes no such thing is possible. And it is this claim with which Sidgwick should more fully engage: his criticisms of perceptual intuitionism are too swift (see CD 6.9).

Sidgwick's assault on ancient ethics continues with a charge of circularity against the Stoics (see Shaver 2008: 214). In its simplest form, this circularity is said to consist in claiming that we should live in accordance with nature, that our nature consists in following the direction of reason, and that reason directs that we should live in accordance with nature. This is indeed an unhelpful position, but Sidgwick provides no reference for it. He then points out that certain Stoics held that reason directs us to live in accordance with virtue. This breaks out of the first circle but might seem to lead to another, similar to but more problematic than that found in Plato and Aristotle. On this version of Stoicism, virtue is knowledge of what is good, and the only good is virtue. As Sidgwick points out, however, the Stoics made room for the idea that certain things were rationally to be preferred even though they were not themselves good. This did not help them, however, since their only recommendations on how we should select these preferable items were that we should choose what is reasonable or choose what is in accordance with nature, either of which brings us back to the original circle.

But even on this conception of Stoicism, it is open to the Stoic to claim that the circle in question is no longer vicious. Imagine that one of the preferable objects is resistance to bodily appetite. According to the Stoic, then, reason requires us to resist, and both our nature and our good consist in following the directions of reason. The view is therefore not practically

empty, and the circle is avoided through reference to reasons which do not depend directly on nature or virtue.³¹

Sidgwick then turns to Butler, accusing him of the original circularity involving reason and nature. Butler claims that it is reasonable to live in accordance with nature, and that conscience is naturally supreme. But conscience is essentially reason, so Butler is telling us that it is reasonable to live in accordance with reason.

The source in Butler of the ‘practical supremacy of Conscience over Self-love’ (200), however, need not, indeed should not, be seen as conscience itself. It is reason construed *independently* of both conscience and self-love that tells us to follow the dictates of conscience.

In the work of real philosophers, then, ‘sham-axioms’ may be implausible; but they are not shams.

3.3 The Self-evident Axioms

Sidgwick’s central statement of the ethical axioms he himself saw as genuine is at 379–84.³² He begins by cautioning us not to expect these principles on their own to tell us what we ought to do in particular cases: they are too abstract for that. In this chapter, Sidgwick argues for essentially three self-evident axioms: ‘Justice, Prudence, and Rational Benevolence’ (382).³³ Let us now examine their various formulations (cf. esp. Schneewind 1977: 291–7).

The first is *Justice (J)*:

J1: [W]hatever action any of us judges to be right for himself, he implicitly judges to be right for all similar persons in similar circumstances. (379)³⁴

31. At 378 n. 1, Sidgwick notes that the Stoics sometimes meant by ‘nature’ what is well adapted to the achievement of certain ‘organic’ ends. But since they believed the universe to be both perfect and entirely determined, they could not base any clear principle for assessing alternative courses of conduct on the observation of actual events. He continues: ‘and . . . their most characteristic practical precepts show a curious conflict between the tendency to accept what was customary as “natural”, and the tendency to reject what seemed arbitrary as unreasonable’. To the first charge, a Stoic might claim that certain events, though unimprovable and hence perfect, might nevertheless be more or less good than certain others. And to make his second charge here stick, Sidgwick would need to show that the Stoics saw custom as arbitrary.
32. Sidgwick also provides an intuitionist argument for hedonism at 400–1 (see CD 3.3).
33. See Langenus 2000: 103. The number is widely contested: see Schneewind 1977: 290 n; also Skelton 2008: 186–202 (note his references in sect. 2). There is also much disagreement, some of which I shall address below, about which statements of Sidgwick’s axioms should be taken as canonical or central.
34. It may be that Sidgwick uses this name for the principle because it seems to him ‘more or less clearly implied in the common notion of “fairness” or “equity”’ (380).

Sidgwick refers us back to an earlier statement:

J1a: If... I judge any action to be right for myself, I implicitly judge it to be right for any other person whose nature and circumstances do not differ from my own in some important respects. (209)³⁵

He sees these as equivalent to:

J2: [I]f a kind of conduct that is right (or wrong) for me is not right (or wrong) for some one else, it must be on the ground of some difference between the two cases, other than the fact that I and he are different persons. (379)

Sidgwick points out that one can construct a version of *J2* that covers how individuals should be treated (one might just replace ‘conduct’ with ‘treatment’). He then claims that these principles have been recognized not in their most abstract and universal forms, but in their application to cases of people related to one another in a similar way. Here we find them in what is known as the ‘Golden Rule’: ‘Do to others as you would have them do to you.’ This rule is imprecise, however, since I might wish you to cooperate with me in wrongdoing, or our circumstances might be relevantly dissimilar. So we require a strict, negative formulation, and because of its strictness and its twice being described as ‘self-evident’ we may take it as the canonical statement of *Justice* (see Schneewind 1977: 293):

J3: [I]t cannot be right for A to treat B in a manner in which it would be wrong for B to treat A, merely on the ground that they are two different individuals, and without there being any difference between the natures or circumstances of the two which can be stated as a reasonable ground for difference of treatment. (380)

Sidgwick is careful to note that *Justice* cannot give ‘complete guidance’. At 209–10, he demonstrates the point by bringing out how the Kantian categorical imperative (‘act as if the maxim of thy action were to become by thy will a universal law of nature’) leads to the collapse of the distinction

35. The sentence before this one, and from which this one is intended to follow, reads: ‘We cannot judge an action to be right for A and wrong for B, unless we can find in the natures or circumstances of the two some difference which we can regard as a reasonable ground for difference in their duties.’ Just before stating *J1a*, Sidgwick claims a disanalogy with scientific explanation, which allows for ‘an accidental or arbitrary element in which we have to acquiesce’. For example, we might explain the presence of more matter in some area than in some other by appealing to antecedent states of affairs. But anyone impressed by the principle of sufficient reason might ask why we should not continue to ask the same question about antecedent states of affairs until we reach something non-arbitrary.

between objective and subjective rightness, once we allow—as we plausibly should—‘that all (or almost all) persons who act conscientiously could sincerely will the maxims on which they act to be universally adopted’.³⁶

In the account of Justice in 3.13.3, Sidgwick next mentions impartiality in the administration of the law as a ‘different application of the principle that individuals in similar conditions should be treated similarly’, noting that this principle is itself implied in the ordinary conception of ‘fairness’ or ‘equity’ (380). He claims also that the principle is acquired through reflection on the similarity of the individuals that constitute a ‘Logical Whole or Genus’, while other, equally important, principles are reached through consideration of a ‘Mathematical or Quantitative Whole’ (380–1).³⁷ One such whole is ‘the good on the whole’ of an individual, and this takes us on to the principle of *Prudence (P)*. Sidgwick notes that the principle of Prudence is sometimes stated as: ‘one ought to aim at one’s own good’. But, he says, this is in danger of tautology, since ‘good’ may be defined as ‘what we ought to aim at’.³⁸ What we can do, however, is refer to an individual’s *good on the whole*, which would give us (though Sidgwick does not explicitly state it as such):

P1: One ought to aim at one’s good on the whole.³⁹

36. The ‘almost all’ must be right. There will be some maxims that really cannot be universalized.

37. LPK 90 and ERRL 330 suggest that a logical whole might be, for example, the series of numbers, understood atemporally, each being a number as much as any other, while a mathematical whole (somewhat confusingly) might be a temporally instantiated series of events, such as a human life.

38. Schneewind (1977: 293) is puzzled by Sidgwick’s point here, since the definition is of ‘good’, not ‘my own good’. Sidgwick’s thought was presumably that one might take ‘my own good’ to be equivalent to ‘good’ in the sense here defined: *my* good is what *I* ought to aim at, and so Prudence would tell me that I ought to aim at what I ought to aim at (see 92 n. 1; Frankena 2000). Sidgwick is then proposing to take ‘one’s own good’ rather as ‘one’s good on the whole’, and indeed that is how he uses the phrase later in the same paragraph. At 391, he speaks of ‘one’s own good on the whole’.

39. At SFEC 44 Sidgwick claims that he failed to supply ‘a sufficient rational justification’ for egoism in earlier editions of the *Methods* which included from the second edition onwards the passage on Prudence we are discussing. This leads to the famous passage on the significance of the distinction between individuals, later introduced into the Concluding Chapter of the *Methods* (498). It might be thought that Sidgwick’s claim implies that he could not have seen himself as offering a rational justification of egoism based on P1 in the *Methods*. But this would be a mistake. Sidgwick felt that P1 alone was insufficient rational justification for egoism, since so many critics failed to agree with him. The passage on the distinction between individuals spells out the assumption on which ‘the rationality of Egoism is based’, an assumption which seems self-evident to Sidgwick (though at SFEC 45 he accepts its denial is a ‘tenable’ position). P1 seems self-evident to Sidgwick because it assumes the significance of the distinction between persons. But it is not *inferred* from that distinction: it involves it. In other words, in the distinction of individuals passage Sidgwick is merely elucidating P1 in the hope that others may find it easier to grasp.

Sidgwick says that he has already referred to this principle, in the following form:

*P*₂: [We should have] impartial concern for all parts of our conscious life. (381)⁴⁰

This can be restated concisely as:

*P*₃: Hereafter *as such* is to be regarded neither less nor more than Now. (381)

Sidgwick makes it clear that all that is meant here is that *mere* difference in time is practically irrelevant; that some good is more certain than another, for example, can still be taken into account:

*P*₄: [T]he mere difference of priority and posteriority in time is not a reasonable ground for having more regard to the consciousness of one moment than [n] to that of another. (381)⁴¹

The form in which the principle presents itself practically to most people is:

*P*₅: A smaller present good is not to be preferred to a greater future good (allowing for difference of certainty). (381)⁴²

Since *P*_{2–5} are all intended as elucidations of *P*₁, and *P*_{2–5} are elliptical, involving temporal neutrality alone and making no reference to the good of the agent in particular or any requirement to promote it, I shall take *P*₁ as the canonical statement of *Prudence*.⁴³

40. Sidgwick refers back here to 124 n. 1.

41. Schneewind (1977: 294) takes this clarification as itself a self-standing principle, and as the canonical form of *Prudence*. I have no strong objection to seeing the claim as an independent principle. But what is problematic is that, like *P*₃ and *P*₅, the clarification makes no reference to the good of the agent in particular. (Like many other commentators, Schneewind fails to state *P*₁, though it is clearly implied in the text. In addition to the principle stated in the note immediately below, which Sidgwick explicitly claims to find self-evident, see also 391–2, where he speaks of the self-evident ‘precept’ ‘to seek... one’s own good on the whole’; and 119, where egoism is said to ‘imply’ the adoption of the agent’s greatest happiness as her ultimate end. Even if my interpretation is correct, I have to admit that Sidgwick should more clearly have distinguished temporal neutrality from egoism.)

42. See 383: ‘I ought not to prefer a present lesser good to a future greater good.’ Note that Sidgwick must be assuming also that I ought not to prefer a smaller future good to a present greater good, which again suggests that he sees many of the statements of his principles as elliptical.

43. See Raphael 1974: 414; Hurka 2014: 144–5. De Lazari-Radek and Singer (2014: 117) allow that Sidgwick takes it as self-evident that one ought to aim at one’s good on the whole, but later distinguish the axiom of prudence from rational egoism (118–19 n. 1; 126). They claim that Nakano–Okuno also believes that rational egoism is not defended as the principle of prudence in 3.13. Nakano–Okuno’s position, however, is not entirely clear. At 2011: 24, she clearly states the principle of ‘rational self-love’ (equated with ‘prudence’ at e.g. 96) as ‘One ought to aim at one’s overall good’, understood as involving a constraint of temporal neutrality (see also 104).

Sidgwick then moves to Rational Benevolence, suggesting that we form the notion of ‘Universal Good’ by comparing and integrating the goods of all sentient beings, just as we form the notion of an individual’s good on the whole by comparing and integrating the different ‘goods’ that occur serially in that individual’s conscious states (382).⁴⁴ This gives us, as a self-evident principle:

B1: The good of any one individual is of no more importance, from the point of view . . . of the Universe, than the good of any other; unless, that is, there are special grounds for believing that more good is likely to be realized in the one case than in the other. (382)

B1 is analogous most straightforwardly to *P2*, which requires impartial concern for all parts of one’s own life.⁴⁵ Just as when we consider a single life as

But at 106, she claims that the principle of self-love is presupposed in utilitarianism. For the view that Sidgwick does not see rational egoism as self-evident, see Shaver 1999: 74–82. Shaver has several arguments for the view that we should understand Sidgwick here to be advocating as self-evident only *P3*—the irrelevance of time—and not *P1* (75–6). (1) Sidgwick says: ‘*All that the principle affirms is that the mere difference of priority and posteriority in time is not a reasonable ground for having more regard to the consciousness of one moment tha[n] to that of another*’ (my italics). But the context makes it clear that Sidgwick is not making a general claim about Prudence here, but contrasting the element of temporal neutrality with a more substantive principle which would forbid one from choosing some nearer good on grounds of its greater certainty. (2) Sidgwick intends the axioms to be clear and distinct, so he would not wish to run together two independent notions in a single axiom. But Sidgwick has already made it clear that understanding a person’s life as a quantitative whole will involve seeing each part, in itself, as important as any other. (3) Sidgwick claims merely that the axiom is ‘implied in’ rational egoism (386). If we include the idea of the individual’s good on the whole, then the axiom ‘virtually implies rational egoism’. But note that at *M3* 388, Sidgwick speaks of Prudence as the ‘principle on which . . . Rational Egoism is based’ (see Schneewind 1977: 306). Why does he change this to the claim about implication? Probably because he had claimed in the SFEC passage that it was the self/other distinction which provided the rational basis. But as I have suggested (CD, chapter 4 n. 39), this is anyway itself involved in *P1*. (4) Sidgwick says several times that consideration of the axioms leads to utilitarianism, and this is puzzling if one of the axioms suggests I should seek my own good. But there is no puzzle if there is no conflict between my own good and the utilitarian end. And, as yet, Sidgwick has identified no such conflict.

44. Irwin claims that Sidgwick argues for utilitarianism from the ‘*apparent* rationality’ of prudence because the argument for utilitarianism depends on our acceptance of the principle that Irwin calls ‘*Extension*’: ‘It is always ultimately rational, and rationally required, to maximize the quantity of good achievable in the largest possible quantitative or logical whole’ (2009: 503; my italics). *Extension*, however, as Sidgwick would admit, could reasonably be denied by the egoist. The argument for utilitarianism does indeed depend on analogy between the temporal neutrality of prudence and the neutrality between individuals of benevolence; but Sidgwick accepts that the analogy is not powerful enough to override the appeal by the egoist to the normative significance of the metaphysical distinction between individuals.
45. Something’s being important to one is equated with one’s having concern for it in 124 n. I agree with Shaver (2014: 623–5) that *B1* can be understood as involving an ‘ought’, against those commentators who see it as concerned only with goodness.

a whole, we see that we should be temporally neutral, so, when we consider all sentient beings as a whole, we see that we should be impartial between different individuals, giving no priority to one over another, except in so far as more good can be promoted. As we saw, P2 is elliptical, stating only one aspect—temporal neutrality—of P1. One could respect P2 by having very little concern for one’s own good at all, as long as that concern was temporally neutral. It is P1 that makes the normative claim that this would be unreasonable: one should promote one’s own good on the whole.

For this reason, it is no surprise to find Sidgwick offering the following principle, immediately after B1:

B2: [A]s a rational being I am bound to aim⁴⁶ at good generally,—so far as it is attainable by my efforts,—not merely at a particular part of it. (382)⁴⁷

Sidgwick then claims that from B1 and B2 we can deduce the abstract principle of Benevolence.⁴⁸ What we might expect, given the analogies with the argument for Prudence, is:

B*: One ought to aim at the universal good.

What we in fact get is:

B3: [E]ach 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, or less certainly knowable or attainable by him. (382)⁴⁹

46. We should not understand Sidgwick to mean by ‘aim’ any kind of conscious direction by the agent. He is talking about ultimate reasons for action, and the notion of the good’s being aimed at is equivalent to that of its being promoted by the action itself. See Hooker 2000a: 356.
47. It might be thought that Sidgwick does not need B2, given his attraction to a ‘fitting-attitude’ view of goodness (see CD 3.1). Consider e.g.: ‘that one ought not to prefer one’s own good to the greater good of another . . . is a purely formal principle, and is evolved immediately out of the notion of “good” or “desirable”, if this notion is used absolutely; as it then must mean “desirable from a universal point of view” or “what all rational beings, as such, ought to aim at realising”’ (MB 27); see Schneewind 1977: 299–300; Shaver 2014: 180–1. I do not read B1 as involving a fitting-attitude conception of good because there would then be no need for B2. One might of course see both B1 and B2 together as spelling out the implications of a fitting-attitude account.
48. The claim of deducibility is introduced only in M5. I take B3 to be, in effect, the conjunction of B1 and B2.
49. Restated as ‘I ought not to prefer my own lesser good to the greater good of another’ at 383. Schneewind suggests that, because Sidgwick derives B3 from B1 and B2, B3 is not self-evident, despite the fact that Sidgwick describes its restatement as self-evident (see also 382). But nothing prevents some principle from being both self-evident and deducible from other principles (whether self-evident or not). As Shaver (1999: 62 n. 1) notes, Sidgwick himself frequently treats the principle of Rational Benevolence as a self-evident intuition (see e.g. xxxv; 388).

On the face of it, B₃ might be seen as close to a restatement of B₁. But it is fairly clear that Sidgwick sees it as equivalent to my B* (see Skelton 2008: 190–1; Irwin 2009: 497). Immediately below he contrasts B₃ with the practical principle that each person ought to promote the good of a limited number of human beings, implying that B₃ requires the promotion of the universal good. At 385, B₃ is restated as:

B₄: The good or welfare of any one individual must as such be an object of rational aim to any other individual no less than his own similar good or welfare.

At 386, we are told that the happiness of others is the ‘aim’ of Rational Benevolence, and at 400 that Rational Benevolence directs us ‘to the pursuit of universal happiness’.

B₃, then, has to be taken as the canonical statement of the principle of benevolence. But there is a strong case for seeing it as equivalent to my B*.

As Schneewind notes (1977: 297), it is striking, given the care with which Sidgwick sets out the four conditions for maximum certainty in 3.II.2, that he never applies them directly to his own axioms. Schneewind continues:

The explanation is that much of the work of testing has already been done, and the rest is to be done in the remainder of the *Methods*. The key terms in each axiom, unlike the mixed descriptive-*cum*-ethical notions involved in many common-sense rules, are purely ethical terms. At their core is a simple notion, which has been carefully distinguished from other notions. To anyone who understands these terms, the notion must now be clear. The non-ethical concepts involved—concepts of action, self, other, earlier, later—are so basic that to give a critical examination of them would require a digression on other branches of philosophy. The axioms are meant to hold with no exceptions and no limitations, and therefore no question of their precision can arise in the way it arises for common-sense rules. Thus the first test has already been passed.

One significant problem, however, is that, though the first test as stated in 3.II requires precision only of the terms in the relevant proposition, it is clear that Sidgwick also requires precision at the level of action-guidance as well. Consider, for example, 228, where Sidgwick notes that practical indeterminacy is as regrettable a form of imprecision in legal as in ethical principles,⁵⁰ and the many cases in which common-sense morality is criticized for its

50. ‘[R]ules of duty ought to admit of precise definition in a universal form.’ Sidgwick then goes on: ‘this assumption naturally belongs to the ordinary or jurial view of Ethics as concerned with a moral code: since we should agree that if obligations are imposed on any one he ought at least to know what they are, and that a law indefinitely drawn must be a bad law’. See CD 1.1.

indeterminacy. J₃ perhaps is sufficiently clear and precise, though it is primarily formal or regulatory, having no major practical implications of its own (see Broad 1930: 223–4). P₁ and B₃, however, are more substantive. Sidgwick admits (379) that they are too abstract to guide action, but even when hedonism is incorporated into them they will run into the problems of indeterminacy in empirical hedonism which Sidgwick himself exposes in 2.3–4. It is hard to deny that Sidgwick should have been stricter on his own principles (see Schultz 1975: 347; Phillips 2011: 103; Hurka 2014: esp. 140).

Schneewind continues: ‘The elimination of non-rational sources of certainty required by the second test is assured by the fact that the axioms use no terms carrying merely conventional, legal, or traditional meanings.’ Schneewind is perhaps assuming that we ourselves can ascertain whether our acceptance of the axioms is merely an ‘impression’ or ‘impulse’ (339). More is required than Schneewind mentions here, but it must surely be clear that Sidgwick reflected hard and long on his principles.

If we take Prudence and Benevolence to be asserting non-exclusive principles, and not normative egoism or utilitarianism, then there is clearly no problem of mutual inconsistency between Sidgwick’s three axioms. (Of course, the inconsistency between egoism and utilitarianism will be a major topic of discussion in the Concluding Chapter.)

What about the fourth condition: non-dissensus? J₃ is likely to be accepted by almost everyone. In effect, it requires that moral judgements be universalizable in the sense that judgements that differ over cases must be justified by reference to some properties of those cases other than the mere identity of any individuals involved.

P₁ is more difficult. Of course egoists will accept it, and dogmatic intuitionists might be prepared to accept it as capturing the virtue of prudence (96). But a utilitarian will almost certainly reject it as unnecessary and misleading. According to the utilitarian, I have reason to promote my own good on the whole, or indeed anyone’s good on the whole, only in so far as I have reason to promote good on the whole—the universal good.

What about B₃? It will be rejected by the egoist, as will both B₁ and B₂, for the same sort of reasons as the utilitarian rejects P₁.⁵¹ The egoist might

51. Shaver (2014: 622) claims that egoists can accept B₁, since they do not take up the point of view of the universe. But taking this point of view merely involves considering all sentient beings as a group. The egoist has nothing to fear from taking such a point of view, and refusing outright to adopt it would be as unreasonable as an egoist with a bias towards the present refusing to consider her life as a whole. Shaver takes the point of view of the universe ‘to be a point of view from which I have no special concern for myself’ (626). But the point of view

be prepared to take the ‘point of view of the universe’—to consider the goods of all collectively. But she will not accept that she should have no more concern for the good of any one individual than for that of any other, or that she has a reason to promote the universal good. What about the dogmatic intuitionist? Even if B₃ is taken not to be equivalent to utilitarianism, but as one principle among others, it is still likely to be rejected on the grounds of its extreme impartiality (Irwin 2009: 512–13). It forbids an agent from giving *any* priority to those close to her—her children, or friends, for example.

Sidgwick notes this aspect of common-sense morality, but explains it on the ground that, in practice, B₃ will recommend that people pay special attention to those close to them. He continues:

I think that a ‘plain man’, in a modern civilised society, if his conscience were fairly brought to consider the hypothetical question, whether it would be morally right for him to seek his own happiness on any occasion if it involved a certain sacrifice of the greater happiness of some other human being,—without any counterbalancing gain to any one else,—would answer unhesitatingly in the negative. (382)

Sidgwick seems to be implying that that the ‘plain man’, on reflection, would accept that the correct principle of benevolence is entirely impartial, and that any priority he gives to those close to him is justified only by its promotion of universal good.⁵² This should come as no surprise. Early

of the universe is clearly meant to be analogous to the point of view of a single individual (several times Sidgwick speaks of the ‘universal point of view’: xx; 405; 414). This point of view involves considering ‘the series of our conscious states’, and on the basis of that consideration ‘*obtaining*’ the maxim of prudence, understood in terms of temporal neutrality. If the point of view of the single individual were itself understood as one involving temporal neutrality, there would be no debate to be had between the temporal neutralist and an opponent who accepts a bias towards the present. Any victory for temporal neutrality would be achieved by fiat, and the same would be true of neutrality across individuals in the case of B₁, *mutatis mutandis*. When Sidgwick says that a utilitarian might be able to persuade an egoist to accept utilitarianism (420–1), it is not merely because the egoist adopts the point of view of the universe but because she claims that her happiness is good from the point of view of the universe. Shaver goes on to argue that in fact an egoist could accept B₂ if one understood ‘from the point of view of the universe’ at the beginning of B₂ (627). Shaver admits that Sidgwick does sometimes treat B₂ as inconsistent with egoism, claiming that ‘given the agreement test, this must be sloppiness’. There is sloppiness in Sidgwick’s argument, but it results from his failure properly to consider the implications of disagreement for his fourth test. (On Shaver’s interpretation, Sidgwick is also of course sloppy in not stating B₂, and indeed B₃, non-elliptically.)

52. Cf. 252–3 (see CD 6.2.2), where Sidgwick claims that, though it is hard to ‘say decidedly’ that the common-sense principle of benevolence is not equivalent to utilitarianism, there is current a ‘lower and narrower estimate of the services that we are held to be strictly bound to render to our fellow-men generally’.

in the *Methods*, he has already told us that seeing one's end as the happiness of, say, one's family 'seems arbitrary, and probably few would maintain it to be reasonable per se, except as the most practicable way of aiming at the general happiness' (10; see CD 1.5). Sidgwick's view is that a dogmatic intuitionist who continues to insist on a non-impartial principle of benevolence has not even reached the level of reflectiveness of a plain man, and so any disagreement she might have with Sidgwick poses no great difficulty to his claim of self-evidence for B₃. As we shall see, this is both uncharitable and implausible.

Note also that Sidgwick moves from the question of whether I should be impartial in my treatment of others to that of whether I should be impartial between myself and others (see Shaver 2014: 190). Perhaps Sidgwick's thought is that, since the plain man will see no justification for prioritizing his own good over the overall good, he will also see no justification for prioritizing the good of those close to him over the impartial good. That may be true in some cases, but it need not be, and we might have expected Sidgwick to recognize the possibility of a self-abnegating form of Christianity which disallowed priority to self but permitted or even required priority to those close to one.⁵³

Further, Sidgwick himself later says, in the passage we have already discussed:

It would be *contrary to Common Sense* to deny that the distinction between any one individual and any other is real and fundamental, and that consequently 'I' am concerned with the quality of my existence as an individual in a sense, fundamentally important, in which I am not concerned with the quality of the existence of other individuals: and this being so, I do not see how it can be proved that this distinction is not to be taken as fundamental in determining the ultimate end of rational action for an individual. (498; my italics; see CD, chapter 4 n. 39)

Perhaps the only way to make the plain man passage consistent with this is to see it as concerning the plain man's views on what is 'morally right' *as opposed to* what is most reasonable. That might explain why Sidgwick thought the plain man would be so 'unhesitatingly' impartial. But anyway it seems unlikely that the plain man would accept that his giving priority to himself would be morally wrong in all cases, even in the narrow sense, given the place of the

53. The substitution of 'in a modern civilised society' for 'in this age and country at least' (M3 382) may be telling.

virtue of prudence within common-sense morality and the limits on what common-sense morality will be thought by the plain man reasonably to demand of any individual (see Parfit 2011: 1.453; Hurka 2014: 147).

None of this matters greatly, however, once we see that a utilitarian will reject P₃ and an egoist will reject B₃. Sidgwick clearly sees both positions as reasonable, and so his fourth condition requires him to suspend judgment on both principles.

Sidgwick's fourth condition for certainty in 3.13.3 was stated in terms of 'universal acceptance' until the fifth edition of the *Methods*, when the condition was stated in terms of non-dissensus rather than consensus (see CD 4.3.1). 3.14.4, however, remained unchanged from the fourth edition. Here Sidgwick claims that his conclusions appear to him 'to be in substantial agreement—in spite of superficial differences—with the doctrines of those moralists who have been most in earnest in seeking among commonly received moral rules for genuine intuitions of the Practical Reason' (384). It is clear, then, that in this section he is looking for agreement with epistemic peers rather than seeking to explain away disagreement with apparent epistemic peers.⁵⁴

Sidgwick has already discussed the history of intuitionism in British moral philosophy in a note appended to 103–4. He suggests there that, because of the opposition to Hobbes's view that morality was grounded merely on the contingencies of self-interest, writers such as Cumberland and Clarke sought a more fundamental and intuitive basis for ethics, but that 'by degrees the attempt to exhibit morality as a body of scientific truth fell into discredit'. It is no surprise, then, to find Sidgwick claiming that the 'earlier intuitional school show[s] . . . a turn of thought on the whole more philosophical than that which the reaction against Hume rendered prevalent' (384), and turning in particular to Clarke, where he finds versions of Justice and Benevolence.⁵⁵

The only other philosopher to receive attention in this section is Kant, who (though not an intuitionist), Sidgwick says, 'is especially noted for his rigour in separating the purely rational element of the moral code' (385–6). Sidgwick again cites the categorical imperative as 'an immediate practical

54. Sidgwick found that the revisions for the fifth edition took longer than he expected: see Sidgwick and Sidgwick 1906: 525.

55. At 384 n. 4, Sidgwick claims that Clarke recognized Prudence only indirectly, citing a passage from Clarke quoted earlier (120) in which Clarke places a limit to the sacrifice that 'Virtue' can reasonably demand from any individual.

corollary' from Justice (see 208–10), and claims that the only ultimate end of virtuous action for Kant is the object of 'Rational Benevolence as commonly conceived'—the happiness of others. Even if this is correct, however, it does not commit Kant to the strict impartiality implicit in Sidgwick's own version of Benevolence.⁵⁶ Further, as Sidgwick notes, Kant disagrees with Prudence. It would be somewhat difficult for Sidgwick to undermine Kant's epistemic capacities regarding Prudence at the same time as appealing to his views on Justice and Benevolence.⁵⁷ But, since he is looking for evidence of agreement rather than explaining away disagreement, he anyway does not try.

56. As Irwin (2009: 518) and Hurka (2014: 149) note, it is anyway more plausible to find in Kant a commitment to benevolence as a *pro tanto* than an overall principle.

57. Further, the note appended to 389–90 roundly criticizes the arguments Kant uses, on Sidgwick's interpretation, to support his version of Benevolence.

5

Virtue

I Intention and Motive

The third book of Sidgwick's *Methods*, 'Intuitionism', is by far the longest, running to over 200 pages. The bulk of it is taken up with discussions of individual virtues and duties, but these are preceded by a generally introductory chapter and a chapter on virtue and duty in general, and followed by a chapter on motives (3.12). Over the last fifty or so years, there has been a good deal of interest in so-called *virtue ethics* (see e.g. Crisp and Slote 1997). Unlike Aristotle, Hume, Nietzsche, and others, Sidgwick is not an authority to whom contemporary virtue ethicists tend to appeal. But it is worth pointing out once again that the *Methods* is Aristotelian in inspiration. Sidgwick saw virtue as a central component of the common-sense morality he was seeking to describe, and the issues he raises about it and the positions he takes on those issues are worthy of the most serious attention even by those hostile to Sidgwick's overall ethical conclusions.

Let me consider first those sections in the introductory chapter on intention and motive (3.1.2–3; see 1.5; CD 2.2). Sidgwick has introduced book 3 with his conception of intuitionism as the view that we have the capacity to see that certain kinds of action are right or reasonable in themselves, independently of at least some of their consequences. In the next section (3.1.2), he considers the objection that he has left out what is really characteristic of intuitionism—its focus on the intentions and motives with which actions are done, rather than 'outward actions as such'.¹

1. At 201 n. 1, Sidgwick rules out 'character' or 'disposition' (and hence also virtue) as the primary object of moral assessment, on the ground that they can be known only by the volitions and feelings they manifest themselves in (see 393–5; CD 5.3.3). Whether he is right in his evaluation is an important question, frequently forgotten in contemporary discussions of virtue ethics (see Crisp 2015a: sect. 5).

Sidgwick responds that this objection rests on a failure to recognize that all moral philosophers will agree that moral judgements primarily concern intentional actions. That is to say:

[W]hat we judge to be ‘wrong’—in the strictest ethical sense—is not any part of the actual effects, as such, of the muscular movements immediately caused by the agent’s volition, but the effects which he foresaw in willing the act; or, more strictly, his volition or choice of realising the effects as foreseen. (201)

At 201 n. 2, Sidgwick repeats his view that negligence is blamed only when there has been past wilful neglect, with a reference back to 60, where he also explains the final clause of the definition above: ‘we do not consider that a man is relieved from moral blame because his wrong intention remains unrealized through external causes’.

I have already discussed this denial of moral luck, and certain other aspects of Sidgwick’s philosophy of action, in the chapter on free will (CD 2.3.1). But let me now note a further problem which arises in the attempt to combine the claim that intention is the focus of moral judgement with a ‘split-level’ account of utilitarianism, according to which there may be good utilitarian reasons for not seeking to put utilitarianism explicitly into practice. Sidgwick later defines utilitarianism as the view that the objectively right action is ‘that which will produce the greatest amount of happiness on the whole’ (411), and is then careful to distinguish utilitarianism as a theory about the ‘criterion of rightness’ from any claim about the end at which we should consciously aim (413):

[I]f 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.

If we interpret utilitarianism in the light of the restriction to intention, we obtain the view that the central claims of utilitarianism would be:

UI (1): An intention is objectively wrong to the extent that it is not directed at the production of the greatest overall happiness.

and

UI (2): An action is objectively wrong to the extent that it is not the result of an intention directed at the production of the greatest overall happiness.

But, as Sidgwick recognizes, it will often produce more happiness overall if agents aim at some end independent of that. So *UI* will often condemn

actions which in fact produce the greatest overall happiness. Something has to give, and I suggest that it should be *UI* rather than utilitarianism as Sidgwick ordinarily understands it. What we see here is another result of Sidgwick's unnecessarily retaining the notions of rightness and wrongness in the form in which they are central to common-sense morality (see CD Preface; I.3). Within common-sense morality, wrongness is conceptually closely tied to blameworthiness, and blame, as an unpleasant sanction, is seen as reasonably applicable only to 'guilty' agents—that is, those who have *intentionally* acted wrongly. If we restrict ourselves to the notion of what is reasonable, we can state the central claim of Sidgwick's utilitarianism (that what we have (strongest) overall reason to do is to produce the greatest overall happiness) without running into difficulties arising out of apparently unfair or unreasonable attributions of blameworthiness to actions.²

It might be objected that this thin conception of reasonableness amounts to nothing more than the kind of evaluation which might be made of, say, an earthquake, purely on the grounds of its consequences. So, just as some earthquake might be bad in producing less than the overall greatest happiness, so some action of mine might be bad for the same reason. Sidgwick might point out, however, the obvious difference between me and an earthquake: I am a rational agent, whose actions are such that there can be reasons for and against them. It is this 'inward' conception of agency that justifies talk of reasons for action in ethics by utilitarians who accept the account of objective rightness or reasonableness one finds in Sidgwick, and would allow Sidgwick to restrict judgements of wrongness to intentional actions (though not to actions *with any particular intention*).

After intentions, Sidgwick moves on to motives (202–3). He claims first that the distinction between motives and intentions is unclear: motives are sometimes seen as that subset of intended consequences which are also desired by the agent, and intentions are usually understood as linked to desired consequences.³ Sidgwick repeats his own preference for a conception of intention that covers all consequences 'foreseen as certain or probable', denying, with a plausible example, the so-called 'doctrine of double effect',

2. See Parfit 2014.

3. At 203 n, Sidgwick notes a further source of confusion for the distinction in our willingness to consider intentions and motives individually or as parts of series. So we may say that morally unlimited ambition is wrong, while allowing that some acts done through such ambition are right; or that some motive, such as a desire for wealth, is in itself acceptable, but bad if the sole motive of a certain politician.

which allows that consequences can be foreseen but unintended (202 n; see Aquinas 1947: II-II. Qu. 64, Art.7). Sidgwick allows, then, for an intention to be wrong (to commit perjury, say) while the motive is good (to save a parent's life).

Is it possible for an act to be right if, though the intention is in accordance with duty, the motive is bad? (If so, this would certainly be a problem for the intention-focused view of ethics Sidgwick subscribes to in the previous paragraph.) Consider, for example, Bentham's case of the person motivated by malice to prosecute someone he believes to be guilty and hence deserving of prosecution. Sidgwick's response to this question is characteristically measured. He agrees that it is certainly our duty to get rid of bad motives if we can, and so a person's intention cannot be 'wholly' right unless she has sought as far as possible to repress any motives known to be bad. But complete suppression of such motives is often impossible, especially if—as in Bentham's case—duty requires us to perform the action in question. And it would be absurd to say that, if we cannot suppress the motive, then we should not perform the action, and implausible to suggest that we should act only on the 'correct' motive. One cannot, in acting for the sake of the object of some desire, 'refuse' to act on that desire.

Sidgwick goes on to consider the Kantian suggestion that moral value correlates with the degree to which an action is performed solely from the motive of doing what is right for its own sake (204–7). He notes first that it will be hard to combine this view with the claim that it is always in a person's own interest to act virtuously, for the reason he mentioned earlier: if the agent believes that claim, she will be unable to exclude it from her motivation. If it is argued that agents should *not* believe that claim, the upshot will be that moral perfection requires ignorance of the relation between happiness and virtue, which Sidgwick finds too paradoxical to accept, advising us that he anyway aims to show later in the book that moral value can accrue to actions performed from some motive other than that of pure duty (see 223; 394–5), and adding that the view is inconsistent with psychological hedonism as well as the doctrine that duties can sometimes be done from self-regarding motives, such as fear of divine punishment. Sidgwick ends the section by pointing out that psychological hedonism, the Kantian view, and the Butlerian position, according to which both conscience and self-love are acceptable motives, are all consistent with intuitionism as he has defined it.

Formal rightness is a matter of motivation by duty. Sidgwick goes on (207–8) to distinguish formal rightness from cases in which the agent believes her action to be right, but is motivated to do it by something other than that apparent rightness. The presence of such a belief constitutes *subjective* rightness, and Sidgwick claims that all intuitionists will agree that an act cannot be ‘absolutely’ right if the agent herself believes it to be wrong.

What about a case in which someone is wrong about what their duty is? Should they perform the action nevertheless? Sidgwick points out that this question is never first-personal in any particular case, since one can never distinguish between what one believes to be right and what is so. But there may be occasions in which we can influence some other person, and here Sidgwick claims that the common-sense view will be that we should not discourage someone from doing what they believe to be right unless what they are proposing to do is especially evil. Whether or not that view is correct matters little, since in ethics we are primarily concerned with objective rightness (that is, what really makes actions right).⁴

Sidgwick’s conclusion on motives at 210 is that they have to be taken into account, and he does this himself towards the end of book 3. The view that what really matters in ethics are motives re-emerges at the beginning of 3.12, after Sidgwick’s pessimistic review of the implications for action of intuitionist morality in the previous chapter. He now distinguishes that sense of motive in which it refers to consequences which are both intended and desired from that in which it refers to desire itself, restricting his own discussion to the latter.

Sidgwick goes on to claim that ethical conflicts between motives are not usually between good and bad motives, but between better and less good (363–4). The only motive which common sense *might* be thought to judge unqualifiedly bad is the desire to harm some other sentient being, and even this turns out to be permissible when it constitutes legitimate resentment. Sidgwick is here making a point close to that at the heart of Aristotelian ethics: that human desires are, in themselves, ethically neutral, virtue consisting in their being appropriate in content, degree, and so on. But in fact he fails to make that point clearly, suggesting instead that most of the other motives commonly criticized morally are ‘seductive’ rather than bad. This

4. Having introduced the notion of objective rightness, Sidgwick then goes on to explain that of universalizability (see 379–80; see CD 4.3.3).

failure, however, does not detract from Sidgwick's main claim in 3.12: that the common-sense view of the moral status of motives runs into the same problems faced by the view of actions, as well as some peculiar to it (365–7). The same problems will arise if we include the moral sentiments (candour, veracity, and so on) in our list of motives, since these sentiments themselves are prompts to actions, about which of course common-sense has shown itself unclear. And if we try to rank motives without clarifying what they require in particular cases, we shall run into deep disagreements on whether, say, veracity trumps benevolence, whether the sense of duty is the only acceptable moral motive, or whether self-love has moral value. Finally, Martineau's suggestion that the only moral impulse is to prefer the higher motive in each individual case is, if descriptive, false, and if prescriptive, both ethically paradoxical and another example of fundamental disagreement.

Further, even if we leave out the moral sentiments and self-love, there will be no reflective agreement on an ethical ranking of motives, except for some vague consensus on, say, the inferiority of bodily appetites to benevolent affections (367–72). And if we attain clarity of ranking in a certain case because motives can combine, we may find ourselves having to choose between a complex consisting of a high and a low motive on the one hand, and a middle-ranking motive on the other. Any ranking (such as Martineau's) is anyway going to seem implausible once we take into account the importance of circumstances in determining our duties. Further, incomplete moral development means that sometimes 'lower' motives, such as resentment, have their place, as, for example, in encouraging the just punishment of wrongdoers (even if in a better world such punishment might result from a 'higher' motive, such as justice). Finally, we must recognize that any conflict of motives should be decided ultimately by the highest: 'the question must inevitably be carried up for decision into the court of whatever motive we regard as supremely regulative' (372). Martineau will of course disagree; but Sidgwick will see that as a result of Martineau's failure to see the need for abstraction and system in ethics.

2 Virtue, Duty, and Supererogation

As I have already suggested, 3.2 is a treasure trove for those interested in virtue and virtue ethics. Sidgwick covers several important topics, and in this section I shall bring out and criticize some of his main claims.

2.1 Objective and Subjective Rightness

Sidgwick begins his discussion of virtue by limiting the term to qualities exhibited in right conduct (219). He notes that we sometimes attribute virtue to actions which are, though subjectively right (that is, believed to be right by the agent), objectively wrong, and even to actions which are both objectively and subjectively wrong (see 394–5). As an example of the former, he mentions generosity, and of the latter, courage. Consider the following three cases of substantial donations to a charity which, though advertising itself as providing assistance to refugees, uses all its profits to supply arms to those promoting some extreme and unpopular cause by acts of terrorism against innocent civilians.

Donation 1: Anna. All evidence available to Anna suggests that the charity is as she takes it to be.

Donation 2: Bryony. All evidence fairly easily available to Bryony suggests that the charity is as she takes it to be. But if she had looked into the matter with greater care, she would have discovered its real nature.

Donation 3: Cara. The evidence fairly easily available to Cara suggests that the charity is not what it seems.

Sidgwick might be read as suggesting that, if we were to speak strictly, we would not call any of these actions virtuous. This seems especially plausible in Donation 3. Though we might describe Cara as ‘well intentioned’, what would dominate our description of her would probably be her irresponsibility and foolishness. Generosity, as Aristotle puts it, requires that we give away our money to the right people, in the right way, and so on, and Cara is clearly not doing that in this case. Rather, she exemplifies the ‘excessive’ vice of wastefulness.⁵

What about Donation 2? This raises the question of how much knowledge we can reasonably expect our imagined virtuous person to possess (see Bailey 2010). Of course, we cannot require her to have perfect knowledge of the consequences of her action. What we can expect, perhaps, is at least a *reasonable* level of care in ensuring that things are as she takes them to be. How much is required will depend on the circumstances of the case. So if Bryony had rather little to go on in forming her view of the charity, we might expect her to have done more to ascertain its real nature, and we might

5. See Aristotle 1894: 1106b21–8; and, on generosity and wastefulness, 4.1.

put her action in the same category as Cara's; whereas if she had a great deal of good evidence that it was bona fide, and would have had to spend many weeks of research to find the truth, we will probably see her action as fully virtuous. But, as Aristotle puts it:

How far and to what extent someone must deviate before becoming blameworthy it is not easy to determine by reason, because nothing perceived by our senses is easily determined; such things are particulars, and judgement about them lies in perception. (1894: 1109b20–3)

As regards Donation 1, however, to describe any admiration we might feel for Anna as 'quasi-moral', to use Sidgwick's term, seems harsh. Surely we have to accept that even the perfectly virtuous agent occasionally makes non-blameworthy mistakes, in the sense that her actions are not always vindicated? In Donation 1, there is no way that Anna could have found out the true nature of the charity she was supporting. Perhaps she even spent a good deal of time researching it. Her own character, then, seems quite untarnished by the deception of others.

It is not clear from the text whether Sidgwick would characterize Anna as lacking 'insight into [the] consequences' of her actions to the point that common sense would see her action as objectively wrong, yet subjectively right. It strikes me as most likely that he intended the lack of insight to be something for which the agent is, at least to some degree, culpable. In other words, Anna's action is both objectively and subjectively wrong. We might, then, see virtue as tied to what the *reasonable* or *epistemically virtuous subject* would believe in the circumstances. So the virtuous (and objectively right) action in any particular case is indeed what the perfectly (epistemically and practically) virtuous agent would do in those circumstances.

What about Sidgwick's suggestion that we attribute some virtues, such as courage, to actions which are both objectively and subjectively wrong? He is quite right. Sometimes people will call bank robbers courageous, even though they accept that the actions in which they display this alleged courage are wrong and that the agents were aware of this. I am inclined to agree with Sidgwick (and indeed Aristotle) that we should restrict the attribution of virtuousness to actions which are right. Bank robbers sometimes show a remarkable, and perhaps indeed admirable, capacity to control their fear. But this is similar to the capacity, say, to swim at high speed. There is nothing *morally* admirable about either capacity.

2.2 The Relation between Virtue and Duty

Sidgwick then (219) moves on to consider whether virtue and duty are co-extensive. He claims that they are to a great extent, but not entirely so. First, we will not say that it is virtuous to perform those duties which most people perform, such as paying debts, giving one's children a decent education, and not allowing one's aged parents to starve. And, second, there are 'acts of high and noble virtue', which go beyond what we usually take to be the agent's duty.

Sidgwick's first point here seems mistaken. Each virtue governs some particular sphere of life, and part of that virtue consists in performing whatever duties are required within that sphere, regardless of whether others are doing so or not. So justice and fidelity require one to pay one's debts, and familial virtues have the same structure. But his second point is more worthy of attention. As Sidgwick puts it: 'Here . . . a difficulty seems to arise; for we should not deny that it is, in some sense, a man's strict duty to do whatever action he judges most excellent, so far as it is in his power.' We shall return to this difficulty. But Sidgwick first asks a further question (220): whether it is as much in a person's power to realize virtue as it is to fulfil duty. His answer is that, in a sense, it is, since we describe as a virtue only a quality of conduct which we think to be *to some extent* immediately attainable by ordinary people. Great musical ability we might describe as a gift or a talent, but not as a virtue. Nevertheless, Sidgwick concludes, one cannot plausibly claim that any ordinary person could exhibit the highest degree of, say, courage, when the opportunity to do so arises. So there is a margin of virtuous conduct which could not strictly be the duty of the individual in question, since it is beyond her power.

Sidgwick's first point here is debatable, and perhaps merely semantic. Consider Aristotle's virtues of magnificence and greatness of soul (1894: 4.2–3). Most ordinary people would not have the resources to possess either of these. Sidgwick might refuse, therefore, to call them virtues. But this appears to amount merely to a stipulation that virtues be restricted to qualities widely attainable; and we might want then to use some other term, such as 'excellences', to refer to the rarer morally admirable qualities.

What of the difference between virtues and talents? Sidgwick objects to Hume's denial of the distinction that it plainly diverges from common sense. It does; but that is not enough to show that Hume was making a mistake.

There is perhaps more to be said for the idea that assessing a person's character is, in a sense, more like assessing someone's musical gifts than assessing whether someone has performed their moral duty or not. But, as I suggested above, it is not the same. Rather, there is a sphere of moral assessment of people's characters the scope of which is broader than what lies within their control and indeed beyond the territory of subjectively right and wrong action (see Adams 1985). This would help explain the phenomenon Sidgwick mentions at the end of this paragraph—that 'we can distinguish a margin of virtuous conduct, which may be beyond the strict duty of any individual as being beyond his power'. We might, that is to say, accept that 'ought' implies 'can', in the sense that no one can be required to show more courage on any occasion than they are capable of showing. Nevertheless, we can *imagine* someone's showing significantly more courage than they are capable of, and we accept that this *would* be praiseworthy and admirable.

2.3 Supererogation

But what about the issue of supererogation (see 243–4; 492–5; CD 7.2.4)? This was not a problem for Aristotle, since the idea of going beyond duty is, within our common-sense morality, a Christian concept. According to Aristotle, one's duty is to do what the virtuous person would do. This is not quite the same as Sidgwick's strict duty to do whatever one judges most excellent, though of course in the case of the virtuous person herself the two duties will coincide. But in neither case is there anything lying *beyond* duty. Sidgwick, however, is working with the common-sense morality of his time, and of his readers. As Sidgwick puts it (220), we cannot accept that a 'truly moral' person might say to herself: 'This is the best thing on the whole for me to do, but yet it is not my duty to do it though it is in my power.' This seems to common sense an 'immoral paradox'. And yet, Sidgwick points out, we do count as virtuous certain acts which are not also counted as duties—such as a rich person's living frugally so that her income can be used for the public good.

Sidgwick's own solution to the problem is to distinguish between what is wrong and what is to be blamed (221). Sometimes the gap is a result of lack of knowledge of the circumstances of other agents. So I may be sure that I ought to donate to some worthy cause; but, since I do not know my neighbour's financial circumstances, I cannot be sure that she is in the same

position morally, and so refrain from blaming her if she does not donate. Often, however, I will not blame her even when I know she is in at least as good a financial position as me. Sidgwick's explanation of such cases is that we think that moral progress will best be achieved through our praising actions well above the common standard of duty (which is set too low) and our blaming actions that fall well below that standard. As he points out, this standard is vague and contingent, and so not useful in moral theory for distinguishing virtue and duty.

If Sidgwick's explanation were correct, then when I refrain from blaming my neighbour even though I know she could contribute, I do so purely for pragmatic reasons. I do think it is her duty to donate, but I refrain from blaming her because I see blame as a scarce resource which is best employed on bringing those below the common standard up to par. As an interpretation of common-sense morality, this seems—for Sidgwick, unusually—off-beam. The reason I do not blame my neighbour is that I firmly believe she has done nothing wrong.

How should a modern virtue ethicist deal with supererogation? One option, of course, would be to accept Sidgwick's rationale, and return to the Aristotelian position in which supererogation plays no theoretical role.⁶ A second option would be allow for different levels of virtue. 'Basic virtue' might be performing one's duty and no more, while 'ideal virtue' would be going beyond duty (see e.g. Dreier 2004: 149).⁷ This is likely to seem unattractive to most virtue ethicists, who will prefer not to describe as virtuous the kind of agent of whom Sidgwick's 'immoral paradox' is true. A final option would be to retain supererogation, and give up the idea that duty is to be defined by reference to what a virtuous person would do. Given the emphasis in modern virtue ethics on such definitions of duty, this may seem too radical a suggestion. But in fact these modern accounts appear unlikely to succeed (Crisp 2015a). Even if one is virtuous it is no reason to do an action that people like oneself characteristically act in that way. Consider a virtuous person saving a drowning child from the pond. The reason she has for doing so, and the one she will adduce, is the need of the child. The fact that people like her characteristically save children in such situations is true,

6. Note that at 243–4, Sidgwick appears readier to accept the common-sense position on supererogation (see CD 6.1). But in that passage, the 'immoral paradox' is not mentioned (or of course resolved). For further discussion and a defence of the Aristotelian position, see Crisp 2013a.

7. An alternative might be to see virtue as correlated with duty, but to allow for supererogatory action which exceeds virtue in value.

but not normatively relevant. By incorporating a notion of duty independent of the virtuous person's own character or choices, one can both allow that the virtuous agent will go beyond duty, and yet not blame the non-virtuous for failing to do so.

3 The Nature of Virtue

3.1 Action and Emotion

At 221–8, Sidgwick begins by noting that we apply the terms 'virtuous', 'just', 'brave', and so on not only to acts, but also to persons. This raises the question whether either application is primary. Sidgwick suggests that, since we do not think of these attributes as belonging to acts 'considered apart from their agents', virtue should be seen primarily as a relatively permanent quality of the mind.

I am inclined to agree with Sidgwick that *virtue* is best understood as a trait of character. But this is not one of the terms that he has noted we apply to both acts and persons. When it comes to the term 'virtuous', should we see its application to persons as in some sense prior to its application to acts? Sidgwick suggests that we do not think of this quality as belonging to an act considered apart from its agent. That seems not quite right. Consider a (non-deviant) case in which I rescue a child from drowning. My action can be recognized as virtuous independently of my agency, in so far as it is the act *that a virtuous person would do* (see Aristotle 1894: 1105b5–7). Indeed, I myself might not be exhibiting any *virtue* at all. I might know that this child has rich parents and be entirely motivated by the possibility of a reward. If I had not known this, I would have left her to drown. It is true that this account of what it is for an action to be virtuous does make reference to a person—the virtuous person. But in fact that is not required, for 'virtuous' here is equivalent to 'right'. What is virtuous is what it is my duty to do (or, in cases of supererogation, perhaps what it is excellent to do), and that explains why it is the action performed by the virtuous person.

So should we reverse Sidgwick's claim, and make virtuousness primarily a property of actions? This seems to me unnecessary, and indeed potentially misleading, since it may conceal the way in which the concepts in this area are related. As reflection on Aristotle's doctrine of the mean makes clear, all we need are the concepts of 'rightness', 'excellence' (i.e. 'supererogatoriness'),

and 'virtue'.⁸ Rightness is of course a property of actions, and if we take it to be equivalent to something like 'appropriateness' it can be attributed easily to feelings and emotions as well. Virtue may then be understood as a disposition to perform right (or excellent) actions and to feel appropriate (or excellent) emotions (in the right way, in relation to the right people, and so on).⁹

Sidgwick himself agrees that our conception of virtue should not exclude emotion (222–3), denying the Kantian view that emotions can have no moral worth. One of his several examples is gratitude. Expressing thanks to a benefactor has some moral worth; indeed, it may be the most important element in a proper response. But expressing thanks while feeling genuinely grateful is yet more valuable. This might appear to take the realm of assessment of virtuous conduct beyond the voluntary, and perhaps beyond the question of blameworthiness. Consider a case in which you give me some gift, and I thank you, but feel no gratitude. Even if I wish it, I cannot will myself to feel grateful. Further, it may not even be my fault that I am unable to feel gratitude—it could be a result of my harsh upbringing. Nevertheless, my state is certainly open to criticism. It seems that this criticism often takes the form of blame, and this might seem unfair, since it is, if expressed, equivalent to punishing the innocent.¹⁰ There are several options open to a proponent of virtue at this point. First, she may wish to restrict the scope of her theory to the voluntary. But if she restricts blame to the voluntary, then she may have to exclude emotions from her account of virtue. Second, she may allow blame to extend beyond the voluntary. But this risks punishing the innocent. Or, third, as I have suggested above, she may recommend that we restrict blame to the voluntary, but allow moral criticism a broader scope, in which it comes closer to aesthetic criticism than paradigmatic blame directed at voluntary actions (see 324–5).

Sidgwick goes on (223; see 205) to make a point which became standard in virtue ethical critiques of Kantian ethics in the work of Bernard Williams and others.¹¹ We are quite prepared to attribute virtue to conduct where

8. See Aristotle 1894: 1106b21–8 (cited at CD, chapter 5 n. 5). Virtue ethicists who wish to expunge the notion of the supererogatory from their account can of course omit the concept of 'excellence'.

9. Can sense be made of the idea of 'supererogatory' feelings? Daniel Star suggested to me the case of feeling compassion for someone when they are not so badly off that a failure to feel compassion would be inappropriate.

10. For the claim that Bernard Williams may be committed to a form of such punishment, though one inflicted through moral criticism other than blame, see Crisp 2013b: 12.

11. The point had often been made before Sidgwick.

there is no thought of duty on the part of the agent, such as saving someone from death on the basis of a spontaneous sympathetic impulse. And in some cases, to use Williams's nice phrase (1981b: 18), such a thought would be 'one too many'. Sidgwick gives the excellent example of humility. A genuinely humble person need not be someone conscious of her fulfilling a duty or exhibiting a virtue through her humility.

3.2 Virtue and Motivation

At 223–4, Sidgwick argues that in the case of certain virtues, we are happy to attribute them quite independently of motivation, whether it be a sense of duty or an emotion. For example, someone has the virtue of veracity if what she says results from a disposition to speak the truth, regardless of whether she does so through regard for virtue, a sense of the 'degradation' of falsehood, a belief that speaking truthfully is in her interest, or sympathy with others. Again, anticipating arguments of Williams (1981c), Sidgwick makes the same point about justice and fidelity.

This raises another important question for virtue ethicists. Are there virtues which are to be understood solely in terms of the actions they produce, regardless of the motivation behind them? As we have seen, in the case of any right action, it can be described as virtuous independently of the actual motivation of the agent: it is what the virtuous person would do in those circumstances. But Sidgwick's claim here is about virtue, not about the virtuousness of action. I myself find most plausible the Aristotelian conception of virtue, according to which the actions it produces must be done in the right way. (I would not want to claim that this is, or is not, the view of common sense: this is a place where common sense is indeterminate.) So a person who habitually tells the truth *merely* because she fears the consequences for her own self-interest of lying, and who would indeed lie were she convinced she could get away with it, seems to me not to have the virtue of veracity. The same is true, *mutatis mutandis*, of a judge who makes just decisions merely because she realizes that this is the best way to advance her career, and who would knowingly convict an innocent person were she to believe that to be in her own interest. We might wish to speak of someone who habitually tells the truth or makes just decisions purely on the basis of self-interest as *quasi-veracious* or *quasi-just*, but it would seem unnatural to ascribe full-blooded virtue to someone who in fact cares nothing about that virtue or its characteristic objects.

Sidgwick notes that common sense often does take motives into account in assessing the *difficulty* of acting morally, so that someone who does the virtuous thing in the face of strong temptation to act otherwise manifests virtue to an especially high degree. But this raises another serious issue for the virtue ethicist (224–5). For we are also inclined to agree with Aristotle's claim that the perfectly virtuous agent will not experience any internal conflict. Sidgwick's solution to the problem is to allow our conception of virtue to include not only the Aristotelian ideal of perfect excellence, but also an element exhibited in the struggle by imperfect agents to attain this ideal:

Thus in proportion as a man comes to like any particular kind of good conduct and to do it without moral effort, we shall not say that his conduct becomes less virtuous but rather more in conformity with a true moral ideal; while at the same time we shall recognise that in this department of his life he has less room to exhibit that other kind of virtue which is manifested in resistance to seductive impulses, and in the energetic striving of the will to get nearer to ideal perfection. (430; see CD 7.2.3)

This eirenic attempt by Sidgwick to make room in our conception of ethics for both the moral saint and the moral hero seems very much in line with any inclination he may have had towards the methods of Aristotelian dialectic (Aristotle 1894: 1145b2–7). He might perhaps have added a point about the practical benefits of praise in this kind of case, analogous to that he made about blame in connection with supererogation. A moral hero is less morally perfect than she might be, and to that extent open to moral criticism—perhaps even to blame. But given that she is struggling to overcome the inclinations she would be better off without, it will be more helpful to praise her for her efforts than to blame her for having the inclinations in the first place.

Sidgwick then inquires further into the intellectual requirements to be met by an agent before her action can be considered virtuous (225–6). Having repeated the point that in certain cases no moral judgement need be passed on the action by the agent (again, Sidgwick uses the example of someone impelled by sympathy courageously to come to another's aid), Sidgwick claims that the action cannot be 'even vaguely thought to be bad'. Here again he would have benefited from a clearer distinction between two conceptions of an action's being virtuous: first, that in which an action is virtuous purely in so far as it is what a virtuous person would do, and, second, that in which an action is virtuous if it is done with the appropriate

emotions, feelings, thoughts, and so on. So someone who does the action that the virtuous person would do in some situation, but who thinks her own action to be quite wrong (Huckleberry Finn comes to mind), is doing the virtuous action in the first sense, but not in the second.

3.3 The Cultivation and Value of Virtue

Having summarized his conclusions so far (226–7), Sidgwick then (227) notes that, even when we cannot immediately realize virtue, we are nevertheless subject to a duty to cultivate it. This duty extends to all virtuous dispositions to the extent that we can increase our tendency to perform the relevant acts in future. Now it could be that such acts are anyway under the control of the will and this raises the question why we might need the virtues. Sidgwick's answer is that the virtue will make it more likely that we shall do the right thing when time is too short for the deliberation which might result in our willing what is right.

One objection to Sidgwick's claim, by someone doubtful of the theoretical importance of the virtues, might be that he underplays the power of the will. There is no reason why willing must be preceded by deliberation. So we can imagine an agent who always wills the right action even in cases where she is called upon suddenly to act. But here the question arises of why we should think that such an agent lacks the relevant virtue. For she appears to have the disposition to act rightly, and that is what a virtue is. Sidgwick should not have drawn such a sharp distinction between virtue and the will.¹²

From the other direction, a virtue ethicist might claim that Sidgwick does not recognize the most important reason for cultivating the virtues. According to Sidgwick, virtue is valuable only in so far as it leads to the right action (and, presumably, the appropriate emotion) (see CD 5.1). Let me quote here an important passage from Sidgwick's discussion of the value of virtue:

From a practical point of view, indeed, I fully recognise the importance of urging that men should aim at an ideal of character, and consider action in its effects on character. But I cannot infer from this that character and its elements—faculties, habits, or dispositions of any kind—are the constituents

12. It is possible that she has such a disposition only in a weak sense, the attribution of the disposition being based merely on what she does. That is, she acts quite by chance, but always rightly. In any real-life case, however, someone who always acts rightly will have the categorical basis of the virtuous disposition. For more on this, see the discussion in the main text.

of Ultimate Good. It seems to me that the opposite is implied in the very conception of a faculty or disposition; it can only be defined as a tendency to act or feel in a certain way under certain conditions; and such a tendency appears to me clearly not valuable in itself but for the acts and feelings in which it takes effect, or for the ulterior consequences of these,—which consequences, again, cannot be regarded as Ultimate Good, so long as they are merely conceived as modifications of faculties, dispositions, etc. (393; see 201 n. 1; 394–5)

If we return to the full paragraph on 227, we can see it as a good example of Sidgwick's allowing his own hedonistic and consequentialist views to influence his account of common-sense morality. The relevant paragraph begins by mentioning the duty to cultivate virtue that *we* recognize, and it is natural to read the rest of it as spelling out the common-sensical justification for that duty—that virtue is instrumental to acts morally valuable in themselves. But it seems to me equally in line, perhaps more in line, with common sense to ascribe moral value to the virtues themselves, as constitutive of an admirable ideal of character (see Shaver 2008: 213). In contrast to Sidgwick, consider the views of another intuitionist, W. D. Ross, writing three decades after Sidgwick's death:

The first thing for which I would claim that it is intrinsically good is virtuous disposition and action, i.e. action, or disposition to act, from any one of certain motives, of which at all events the most notable are the desire to do one's duty, the desire to bring into being something that is good, and the desire to give pleasure or save pain to others. It seems clear that we regard all such actions and dispositions as having value in themselves apart from any consequence. And if any one is inclined to doubt this and to think that, say, pleasure alone is intrinsically good, it seems to me enough to ask the question whether, of two states of the universe holding equal amounts of pleasure, we should really think no better of one in which the actions and dispositions of all the persons in it were thoroughly virtuous than of one in which they were highly vicious. To this there can be only one answer. (Ross 1930: 134)

Note again Ross's appeal to what *we* regard as valuable. It seems to me that common-sense morality is indeterminate on this question, and the virtue ethicist has to decide for herself whether to attribute non-instrumental value to a disposition rather than to its manifestations alone. Almost certainly what led Sidgwick to his position was his prior commitment to hedonism. Any disposition can be valuable only to the extent that it produces pleasure, and then only instrumentally. Ross, on the other hand, is both more inclined to pluralism, and also more sympathetic than Sidgwick

to the Aristotelian conception of virtue. Anyone looking at virtue from the Aristotelian perspective will not see it as a mere 'black box', its value to be judged in terms of its results. Virtue is not to be understood just in terms of subjunctive conditionals. It is indeed true of the just person, say, that she *would* act justly in the relevant circumstances. But the reason for that is that she possesses the *categorical basis* of that disposition, which, according to Aristotle, consists in a properly trained character—a pattern of concerns, motives, wishes, desires, and so on—and a perceptual capacity to grasp the salience and normative weight of morally relevant considerations. If anything, I suspect that, once confronted with the choice between Sidgwickian reductionist hedonism about virtue on the one hand and the Rossian view of virtue as good in itself, most people would side with Ross. But that of course does not mean that Sidgwick is mistaken.

The third and final section of 3.2 raises some further deep and important Aristotelian questions about the nature of moral principles and their relation to ethical judgement. It will be best to consider Sidgwick's position here after examining his discussion and criticism of common-sense morality (see CD 6.9).

6

The Virtues

In 3.3–10, Sidgwick begins in earnest his discussion of common-sense morality. This is the first element of a three-part strategy, and is intended to consist in an impartial survey of the ethics of his day. In 3.11.3–9, Sidgwick returns to the virtues and principles he has surveyed and holds them up to the light. From the point of view of self-evidence, of course, he finds them lacking. But they can serve as helpful secondary principles, underpinned by the more fundamental principle of utilitarianism, as he tries to show in his third and final engagement with common-sense morality, in 4.3. In this chapter, I shall proceed by examining Sidgwick's discussions of each virtue serially, running together the survey, the 3.11 critique, and the discussion in 4.3.¹ This is partly for ease of reading, since there is a fair amount of repetition in 3.11. But it is justified also by Sidgwick's own failure to keep criticism out of the survey (see e.g. 302; 313; 316; 322; 324; 355; 355–6; CD 1.1).

I Wisdom and Self-control

One of my central claims in this book is that, because of his desire for a comprehensively action-guiding ethical theory and his consequent antipathy to perceptual intuitionism, Sidgwick failed fully to recognize the role that a well-worked-out conception of practical judgement might play in a developed form of dogmatic intuitionism such as Aristotle's (CD 6.9).² This shortcoming is especially in evidence in 3.3, the short chapter on wisdom

1. 4.3 includes no further discussion of liberality (the subject of 3.8), or courage or humility (covered in 3.10).
2. The title of 3.3 was changed from 'The Intellectual Virtues' in M3. The title in the Contents was never changed.

which initiates Sidgwick's detailed discussion of the virtues and duties of common-sense morality. Though the chapter is in many respects acute and insightful, Sidgwick's views that his three methods are hidden within common-sense morality and that the sphere of virtue is bounded by the scope of currently voluntary action, along with his doubts about the significance of practical judgement, lead him first to interpret wisdom over-systematically and then to distort it into a form of self-control.

The first section (231–3) offers an account of the common-sense conception of wisdom. Sidgwick rightly notes that, in contemporary usage, the sphere of operation of the virtue is seen as primarily practical rather than theoretical, and even when an intellectual is described as 'wise' it will be because of traits that tend to lead to sound practical conclusions, such as breadth of vision, impartiality, and good judgement. Like Aristotle (1894: 1144a23–9), Sidgwick sees as constitutive of practical wisdom a general capacity to discern the right means to common human ends *as well as* the right ends themselves (see 392–3; Whewell 1845: 1.157; 220).³ We would describe an 'accomplished swindler' not so much as 'wise' but as 'clever'. The means–end component is acquired, Sidgwick suggests, as in the case of purely technical skills, through both the learning of certain rules and the development of an 'instinct', dependent on natural ability to an extent but importantly requiring 'exercise and imitation'. This account is similar to that of Aristotle concerning the acquisition of virtue of character in general (1894: 2.1–4), and it is not unreasonable to believe that Sidgwick would have accepted Aristotle's view that the kind of capacity for judgement that enables the wise person to grasp the right ends—both in general and in particular cases—is also acquired through experience:

[T]hough the young become proficient in geometry and mathematics, and wise in matters like these, they do not seem to become practically wise. The reason is that practical wisdom is concerned also with particular facts, and particulars come to be known from experience; and a young person is not experienced, since experience takes a long time to produce. (1894: 1142a12–16)

Sidgwick now turns to a 'subtle question'. If wisdom involves correct judgement about ends, then anyone who holds some single end to be right will refuse to attribute wisdom to any other person who accepts some other end

3. The short discussion of wisdom in 430 notes that if wisdom consists in choosing the right means to the right end, that end can of course be understood in utilitarian terms.

or ends. It might be thought that common sense, because it sees wisdom itself as a quality of goodness or rightness in conduct, will see the end of wisdom itself as goodness or rightness in conduct *rather than* pleasure for himself or any other end beyond such goodness and rightness. Though this thought involves a non sequitur (since there is no inconsistency in believing that wisdom is a quality of goodness or rightness in conduct and that it aims at ends beyond such goodness or rightness), Sidgwick does not trouble to point it out. Rather, he brings in again his notion that common sense, recognizing not only the ends of rightness and goodness as conceived of within dogmatic intuitionism, but also those of egoistic and universalistic hedonism, is inclined to see these ends as in harmony with one another. The wise person, then, will be the person who achieves all three ends. But if those ends are in fact not in harmony (as Sidgwick thinks they are not), then there is a conflict here within the common-sense conception of wisdom which common sense, because of its assumption of harmony, cannot resolve.

Without committing itself to the non sequitur, however, it is open to common sense to hold, with Aristotle (1894: 6.12–13), that practical wisdom is closely tied to the other virtues of character, and that the notions of the agent's own happiness and of the happiness of all will find their place within the common-sense conceptions of the virtues of prudence and benevolence, respectively. So the three-way conflict between methods that Sidgwick sees within common-sense morality can be avoided. There will of course be practical conflicts and dilemmas, in which the happiness of the agent, the happiness of others, and other considerations relating to other virtues, such as those of justice or veracity, have to be weighed against one another in the wise person's deliberation. But it is at exactly this point that her good judgment (her '*instinct*') comes into play, allowing her to make the correct, or at least a justifiable, decision in the light of 'the particulars'.

Another aspect of Aristotelian ethics relevant here is his limiting of the scope of ethics to the voluntary:

Since virtue is to do with feelings and actions, and since voluntary feelings and actions are praised and blamed, while the involuntary ones are pardoned and occasionally even pitied, presumably anyone considering virtue must determine the limits of the voluntary and the involuntary. (1894: 1109b30–4)

As we have seen, Sidgwick feels a similar temptation to link the concepts of rightness and wrongness with the voluntary (1 (see CD 1.5); 59–60 (see CD 2.3.1; 5.3.1)). One obvious question to ask Aristotle is how feelings can be

voluntary. He attempts somewhat indirectly to deal with this question by pushing back the point of responsibility (1894: 3.5): I am blameworthy for being prone to excessive anger, for example, because of my giving in to anger in the past. This prompts one to ask whether what I am really blameworthy for is that past failure rather than my present anger, as well as further questions about the sources of character in one's environment rather than in one's own voluntary choices. Nor does Aristotle reflect sufficiently carefully on the fact that we praise and blame people for virtues and vices independently of any knowledge we might have of how they acquired those traits. This applies to intellectual virtues as much as to virtues of character. A good person will be praised for their practical wisdom because it is in itself an admirable trait, not in so far as they have voluntarily taken steps to acquire it. Aristotle's own account of the virtues, however, is not seriously affected by the excessive importance he gives to the voluntary when it is in his sights; he probably just assumes that, as the virtues of character must be acquired through voluntary action, and these virtues themselves provide the foundation for practical wisdom, then practical wisdom is also within the sphere of the voluntary and hence praiseworthy. But it is not a topic he discusses. Sidgwick, however, reflects at greater length on the role and nature of the voluntary, and it is odd that he does not see the limiting implications of the notion of moral sainthood for the voluntary (224–5; CD 5.3.2). One possibility is that he might have categorized non-blame-involving criticism of others as not part of 'morality proper', but as a matter of 'good taste and refined feeling' (259).⁴ But this would have required from him a much deeper analysis of the nature of blame, for on the face of it, the kind of moral criticism often directed at traits of character, independently of whether the agent is or is not responsible for them, can involve much the same kind of anger, indignation, and hostility as blame.

Focusing on the voluntary, then, Sidgwick starts his own account of practical wisdom, at 233–6, by asking whether it is 'attainable at will, and so, according to our definition, a Virtue'. He accepts that at first sight it might seem not to be so attainable, but then suggests that wisdom's dependence on the presence of certain desires and emotions explains why it is seen as virtuous. Here he might of course have questioned the voluntariness of those desires and emotions, but chooses instead to mention the large disagreements between common-sense moralists about what they should be ('ardent

4. See also his distinction between 'virtues' and 'excellences' at 236; see CD, chapter 7 n. 6.

aspiration' or 'tranquillity of feeling', for example). But there is consensus that *certain* violent passions and appetites are likely to blur moral judgement, and so we might call someone 'voluntarily wise' if they try to restrict their influence. Practical wisdom, then, because of Sidgwick's focus on synchronic voluntariness, is morphing into what Aristotle (1894: 7.1) would call 'continence' (*enkrateia*) (which is anyway something that a truly virtuous person, on Aristotle's view, will not need, not being subject in the first place to the sorts of passions that need controlling).⁵

Sidgwick then considers whether practical wisdom is exhibited not just in grasping the right ends, but also in actually choosing to pursue them. He concludes that it is, since failure to do so, either deliberately or impulsively, is a form of 'culpable unwisdom', again due to lack of self-control. Interestingly, in a comment he makes on cases in which the agent is overcome by impulse so suddenly that she has no chance to deliberate, Sidgwick says:

[I]n this case the self-control or firmness required to prevent unreasonable action seems to be not attainable at will, when it is most wanted. We can, however, cultivate this important habit by graving our resolves deeper in the moments of deliberation that continually intervene among the moments of impulsive action. (235)

Sidgwick here resists the Aristotelian temptation to base current responsibility for actions resulting from character traits on the voluntary acquisition of those traits in the past. For him, even more than for Aristotle, what matters morally is whether the agent could *here and now* have acted otherwise than she did.

Sidgwick's attempt to reinterpret practical wisdom as self-control continues into the final section of the chapter, where he refuses the title of virtue to various other intellectual capacities, such as sagacity (seeing what really matters), acuteness (spotting partially hidden aids or obstacles), ingenuity (devising complicated means), and caution (in the sense of taking material circumstances properly into account).⁶ Sidgwick *is* prepared to call these

5. See 344. Illuminatingly, this conception of wisdom is entirely absent from the beginning of the following chapter (238) and the 'third-stage' discussion of wisdom at 430, where Sidgwick returns to the idea of wisdom as concerned with the choice of ends.

6. At 236 n, Sidgwick allows for another sense of 'caution', as an inclination to take the more certain and less dangerous means to one's ends. He continues: 'In this sense, in so far as the chance in each case of winning the end, and the value of the end as compared with other ends, and as weighed against the detriment which its pursuit may entail, can be precisely estimated, the limits of the duty of Caution may obviously be determined without difficulty.' This appears to suggest that we have a moral duty to exercise caution; but the scope of 'can' here is not clear. Sidgwick may mean that I have a duty to exercise caution so far as I am able, or—at the other

traits ‘excellences’, but they are not virtues because they are not attainable at will. The only further intellectual virtues he allows are caution in another sense (the capacity to continue deliberation in the face of an impulse to act immediately) and decision (the opposite of caution so understood—that is, the capacity to act despite a temptation to continue deliberation unnecessarily).

The chapter ends with a cautionary note concerning how we should understand the ideally virtuous agent. We should accept that perfect judgement, combined with perfect self-control and any appropriate emotional responses, will result in the performance of duty. But we should recognize also that there may be different degrees of intensity in efforts to do what is right. So we might praise someone as energetic if they will what is right with special enthusiasm, while another person, who displays the relevant moral emotions especially strongly, might be described as showing zeal or moral ardour. The breadth of Sidgwick’s moral psychology here (see 39–40; CD 1.3) seems especially striking when set against the extreme narrowness of the conception of practical wisdom developed in the preceding pages.

The discussion of wisdom at 343–5 begins with what is essentially a foreshadowing of Sidgwick’s argument against the ‘sham-axiom’ of temperance at 375, and again seems to be somewhat uncharitable (see CD 4.3.2). Sidgwick claims that if temperance requires that reason never give way to passion, it is either empty (being equivalent to the claim that it is our duty to do what we judge to be our duty) or questionable (if it is alleging that rational reflection is always preferable to unreflectively acting on the basis of instinct or affection). In a note attached to his elucidation of the first

extreme—that I have a duty to make as precise an estimate of the relevant considerations as is humanly possible in the circumstances. The fact that he speaks of the duty of caution in a context in which he so clearly has before his mind the link between virtue and voluntariness might suggest the first interpretation. It might then seem unclear why he is unprepared to extend this account to the other intellectual virtues mentioned in section 3, so that I have a duty to be as sagacious as I can be, as acute as I can be, and so on. The reason he will not do so, of course, is that these virtues require one to reach a certain standard: I could not be called sagacious if I am all but incapable of prioritizing what matters. But this point applies as readily to caution as understood in this footnote. If anything, we see here a move beyond the confines of the voluntary: if I am incautious, that in itself is a violation of duty and hence blameworthy, or at least open to criticism, even if the failure is involuntary. The conceptual apparatus for such evaluation is itself implicit in Sidgwick’s distinction between ‘virtue’ and ‘excellence’ (we may wish not to blame someone for some trait, whether of character or intellect, if they possess it involuntarily, but we can certainly adopt an attitude hostile to it, just as we can adopt a pro-attitude to excellences for which the agent herself bears no responsibility).

conception (344 n. 2), Sidgwick reminds us of the distinction between subjective and objective rightness. If I do what I judge to be right, then my action is subjectively right; but if my judgement is mistaken, then my action will be objectively wrong.

There is, however, an alternative way of understanding temperance or self-control, which, though closer to Sidgwick's first conception, is not empty. Sidgwick's statement of the principle of temperance—'never give way to passion'—is a statement of an Aristotelian 'excess'. The temperate person is the one who will not give way to passion *when it would be inappropriate to do so*. (So—to refer to the second conception—if spontaneity is what is required, then the virtue of temperance will be expressed through giving way to passion in those circumstances.) An account of temperance which involves appropriateness can itself be understood in two ways. In the first, more Aristotelian way, the temperate person is the 'moral saint', who will act rationally, reasonably, and morally without experiencing irrational desire at all. In the second way, which seems closer to what Sidgwick has in mind, the temperate person is a 'moral hero', who will indeed experience desires which though irrational are natural to human beings, and will resist them and act rationally.

Neither of these versions of the 'appropriateness' account is equivalent to the view that we have a duty to do what we believe to be our duty. Someone doing what is clearly irrational, unreasonable, or wrong lacks the virtue of temperance on either account, even if she believes what she is doing to be right. And this is indeed the view of virtue in general which Sidgwick recommends when he limits virtue to qualities exhibited in right conduct (219; see CD 5.2.1). The bank robber who is not distracted from her task by irrational desires, either because she does not have them or because she resists them, is not temperate. In other words, as far as the distinction between objective and subjective rightness goes, the defender of temperance will claim that it is objectively right to do what one believes to be objectively right only when that belief is correct. Whether it is blameworthy to act against one's view of rightness is a question that arises in connection with an entirely different virtue: conscientiousness.

Nor, of course, is either account recommending anything as regards rational reflection itself or any habit of such reflection. On the appropriateness account of temperance, virtue does indeed lie in acting rationally (as will be true of any virtue); but understanding the content of the virtue requires reference to the absence of, or resistance to, irrational desires.

2 Benevolence

2.1 Feeling, Object, and Scope

The first non-intellectual virtue Sidgwick discusses is benevolence, since many (and not just utilitarians) ‘in modern times’⁷ have seen this virtue as supreme and architectonic (238). Benevolence is, then, a rival in this respect to practical wisdom in its broader sense, and it is noteworthy that Sidgwick goes on to allow that affection—though it cannot be a strict duty, since it cannot be summoned at will—does play a role in ‘our common notion of Charity or Philanthropy, regarded as a *Virtue*’ (239; my italics; see 244). Common sense, then, does not identify love of humanity as a mere *excellence*, and—though Sidgwick does not say this—there is no doubt that people blame one another for a failure to feel appropriate affection, independently of any thought about whether such feelings are or are not in our power. There is an opportunity here for Sidgwick to tease out another major tension in common-sense morality, but he fails to take it because, as we have seen, the same tension is present in his own thought (1 (see CD 1.4); 59–60 (see CD 2.3.1); 222–3 (see CD 5.3.1; 6.1)). He satisfies himself with the claim that the only duty in play here can be to cultivate the affection. But even here there are important limits (244–5). Sidgwick notes that love, though it involves a desire for the good of the beloved, is primarily a pleasurable feeling depending on a ‘sense of union’ with the other and a desire for their company which may on occasion override the desire for their good. Sidgwick claims that we *blame* the affection in such cases.⁸ Indeed, common sense will probably not regard *individualized* affection as an essential part of moral perfection. We need to cultivate such affection only to the point required to facilitate the performance of our duties (understood independently from the affection itself).

What does benevolence promote (239–40)? According to utilitarians, benevolence is a matter of promoting happiness.⁹ But others hold that the good consists in perfection, and in particular virtue (and there is no reason

7. By this Sidgwick probably means ‘since Hobbes’ (OHE 159).

8. At 241, Sidgwick notes that we will call ‘weak and foolish’ the affection of those who provide some longed-for object for their beloved which they know will be overall bad for the beloved in hedonistic terms.

9. At 400, however, Sidgwick allows that the principle of Rational Benevolence may require us to promote ideal goods.

to think that we cannot promote the virtue of others by putting them in a position to act virtuously). Sidgwick is cautious, however:

[I]n the case of intense individual affection, the friend or lover generally longs that the beloved should be excellent and admirable as well as happy: perhaps, however, this is because love involves preference, and the lover desires that the beloved should be really worthy of preference as well as actually preferred by him, as otherwise there is a conflict between Love and Reason. (240)

This passage might suggest that Sidgwick believed ordinary people to be ‘unconsciously hedonistic’ not only in their morality, but also in their conception of the good. So though I might ‘consciously’ think that I am encouraging my beloved in some aesthetic project because success in such projects is a constituent of happiness, in fact my concern is to avoid the kind of cognitive dissonance to which Sidgwick refers, in which I find I prefer someone who is not rationally preferable. But as we shall see (CD 7.2.3), this interpretation of common-sense thought leaves it in danger of a different form of dissonance: once the utilitarian view becomes conscious, it will be in tension with my view of rational preferability. But Sidgwick is anyway over-intellectualizing at this point. He admits that many philosophers see perfection as a constituent of good; there seems no reason why ordinary people should not take the same view (see CD 1.5).¹⁰ Indeed, he immediately goes on to recognize this (240), noting that the result is that the common-sense conception of benevolence is unclear on the nature of the good to be promoted. But this unclarity is, of course, common to philosophy also, and not to be avoided merely through attachment to one or other view of the good. Though Sidgwick may indeed be entitled to restrict his own conception of the object of benevolence to happiness, the common-sense thinker is equally entitled to a different, perhaps entirely non-hedonistic, conception. Here, then, we have another example of Sidgwick’s ‘correcting’ the common-sense morality he is meant to be describing impartially (see CD 1.1).¹¹

As far as the *scope* of benevolence is concerned, Sidgwick plausibly suggests that common sense lines up with utilitarianism against those intuitionists

10. At 430–1, Sidgwick claims that, since the main element of perfectionist good is virtue, if virtue itself promotes happiness there will be no conflict between the promotion of perfectionist and hedonist good. But conflict is of course still possible, even if the non-moral elements are less significant than virtue.

11. Note the reference to ‘*Rational Benevolence*’ here (my italics).

who argue that any duty of benevolence to non-humans is merely indirect, in so far as such benevolence assists in the cultivation of benevolent dispositions towards human beings (241; 431). But utilitarianism's commitment to indirectness at another point marks an essential point of difference between its conception and that of common sense (241–2). Utilitarians and common-sense moralists might agree that I should visit some friend in hospital. But, according to utilitarianism, my reason for doing so is that the visit will be part of a pattern of distribution of kindness by me which will promote the greatest happiness overall. There are no direct duties 'to those who stand in special relations to us'. The question for intuitionism is whether it can provide self-evident principles of sufficient precision and clarity to guide us whenever there are conflicts of such duties (imagine, for example, that just as I am leaving for the hospital my mother calls and asks me for some urgent assistance with her tax return).

Before moving on to discuss these special relations, Sidgwick pauses to reflect further upon the relation of benevolence to justice (242–3). He notes that, in a narrow sense, benevolence is sometimes seen as secondary to justice, as beginning where justice ends; but since it is often thought that there are *duties* of benevolence to particular people, it may seem that benevolence and justice are bound up with one another. It is also said that the 'duties of affection' are vague, whereas those of justice are precise. As a generalization, this may well be correct: exactly how much I should love my mother is unclear, whereas I know exactly how much I owe the bank. But duties of affection can sometimes conflict, and in such cases we shall need to make them as precise as we can. All such duties, then, as chapter 3.5 will argue, can be seen as matters of justice, though they can be discussed appropriately under the heading of benevolence without any difficulty.

What about supererogatory benevolence (243–4)? Sidgwick here makes no reference to the 'immoral paradox' or his solution to it, according to which the common-sense view is that apparently supererogatory action is in fact required by duty, though it should not be blamed (220–1; see CD 5.2.3). Rather, he here appears to allow that common sense does admit the possibility of genuine supererogation, seeing it as clearly virtuous, as also 'natural kindness . . . towards human beings generally', though there is something of a tension, Sidgwick nicely notes (244 n), in common-sense admiration for, on the one hand, the careful calculation required for producing the most overall happiness, and, on the other, spontaneous response to the need of others. Further, because such kindness is supererogatory, and because,

as we saw, common sense does not see individualized affection as morally required at all, the question for this chapter is whether common sense can provide clear principles to guide us in kind action rather than in kind feeling (245–6).

2.2 Special Relations

According to Sidgwick, it is quite easy to list the categories of human beings to whom, according to common sense, we have duties of benevolence (246–7): our parents, spouse, children, and, to a lesser degree, our kin; beneficiaries and friends; neighbours and fellow citizens; those of our own race; and in general human beings ‘in proportion to their affinity to ourselves’; our country, and smaller institutions we are involved in. Indeed, we may be required to benefit any human being we come into contact with in certain small ways, and those in urgent need have a special claim on us.

Sidgwick’s mention of the standard racist conception of partiality of his day is evidence of the point that he goes on to make: that the difficulty of making these duties more precise is compounded by recognition of the great variation over time and space in views of the scope and stringency of duties of benevolence.¹² Sidgwick notes the significance placed on hospitality in the past, and the emergence in his own time of the view that a childless person is not bound to leave his estate to collateral relatives. But the racism might also give rise to a challenge to those who uphold non-racist principles of partiality, based on consanguinity or nationality, for example, to demonstrate the difference between the property they see as morally relevant and that of race.

Sidgwick will not accept the view that, since custom and law are analogous, and we accept the general principle that one should obey the law of the country in which one finds oneself, one should follow established customs of benevolence (247). Progressive communities will have some system for repeal of bad laws; but in the case of custom, the only recourse is individual disobedience. Sidgwick then goes on to make a point revealing of the standards he sets for self-evident moral principles. It might be claimed that

12. See Schultz 2004: 647. For an excellent discussion of Sidgwick’s attitude to race, see Schultz 2004: ch. 7. Note that at 10 Sidgwick says that limiting the scope of benevolence to race (or family or nation) seems ‘arbitrary’, and that ‘probably few would maintain it to be reasonable *per se*, except as the most practicable way of aiming at the general happiness, or of indirectly securing one’s own’. Perhaps the best way to attain consistency here is through the distinction between unreflective and reflective common sense.

the principle in play here is that customs should be obeyed until their effects are bad to some extent. But we shall then need another principle, such as utilitarianism, to tell us whether some custom has or has not reached that point, and failure of the original principle to determine its own limits is enough to prove its lack of independence and self-evidence.

Sidgwick distinguishes four ‘heads’ under which to place the duties of benevolence:

1. Involuntary duties (e.g. those based on kin).
2. Voluntary duties (e.g. those based on friendship).
3. Duties of gratitude.
4. Duties of pity (i.e. duties arising out of special need).

These categories, Sidgwick suggests, are really just for ease of discussion, since they cut across one another: children, for example, are often said to have (involuntary) duties of gratitude to their parents. But in this case the duty is not one *based on* the involuntary biological relation in which children stand to their parents. If we see Sidgwick’s categories as an attempt to capture the properties that ground or, as he puts it, ‘rationalize’ the duty in question, their boundaries may seem clearer. There are duties based on biological and other unchosen social relations; duties based on voluntary social relations; duties to benefactors; and duties to potential beneficiaries in urgent need of assistance.

Sidgwick begins his discussion with the first category: duties based on involuntary relations (248–54). From the very start of that discussion, it becomes clear that central to Sidgwick’s strategy is the exposure of disagreement and unclarity within common-sense morality. As far as the child–parent relation is concerned, most would say that mere consanguinity gives parents a claim on their children’s benevolence, but it would be hard for them to agree on the force of that claim, and exactly what is called for from children (248–9). As far as duties of parents to children are concerned, it might be thought that these follow from the universal duty not to harm others, given that the parent causes the child to exist and so is an indirect cause of the suffering it would experience if neglected (249–50).¹³ But this

13. Sidgwick appears unbothered by the idea of harming someone by causing their existence, though of course it could not plausibly be said that the person would be better off had that harmful action been omitted (since they would not exist). On this debate, see e.g. Parfit 1984: App. G; Holtug 2001. I believe Sidgwick is right to be unbothered, since in this context what he *means* by ‘harming’ is causing someone to exist who then suffers, not making worse off someone who already exists. This issue is another good example of how use of superfluous

suggestion passes over the fact that we blame people who hand over their children entirely to others, even if they are well looked after. Parents have a duty to care for their children themselves to some degree; but exactly how much care they are required to give beyond some minimum is not clear. Sidgwick, who had no children of his own, rightly notes that we do not approve of a parent who makes some huge sacrifice to produce a small benefit for their offspring (see 431–2). Nor is there agreement even on whether children have a right to inherit. Further, this vagueness increases as kinship becomes more remote (250). So some think cousins, for example, have a right to inheritance in the absence of closer heirs, while others believe ‘that an unpromising cousin should not be preferred to a promising stranger’.

Mere geographical juxtaposition, Sidgwick suggests, no one would count as a ground for duties (250–1). Rather, we find that the sense of duty to humanity as a whole is strengthened by association of neighbours with one another. The claim about juxtaposition does seem plausible, and common sense might even be brought to a recognition of it through reflection on Sidgwick’s principle of justice as universalizability: just as identity is morally irrelevant, so is spatial position. But it is hard to see why Sidgwick feels the need to shoehorn communal bonds into the category of a duty to humanity at large, and indeed to give the same analysis of at least aspects of the duties of kinship and fellow-citizenship. Such commonalities are seen by common sense as at least potentially morally significant in themselves.

Sidgwick then returns to the theme of unclarity (251–4). It is not clear how far individuals are required to perform certain unpaid functions of government, or voluntary but paid military service—that is, to be patriotic. Nor is there consensus on our general duty to humanity. As we have seen, Sidgwick believes that pure impartial benevolence may be what is required by common sense (349 (see CD 4.3.3); 382 (see CD 4.3.3)), but he notes here the existence of the less demanding view that we are required to avoid causing undeserved harm, and to make sacrifices to help others if those sacrifices are small in comparison to the benefits provided.¹⁴ Sidgwick then considers what he calls the duty of ‘reverence’, that is, ‘the recognition of superiority or worth in others’, seeing the feeling as analogous to that of

concepts, such as the everyday concepts of ‘benefiting’ and ‘harming’, can create unnecessary philosophical problems.

14. Sidgwick does not note here the potentially highly demanding implications of this common-sense principle; but he does so below at 262 (see CD 6.3.3).

love for another. Again, we find much disagreement about whether, for example, one is required to respect rank or social position. The same goes for loyalty. Some believe that some degree of affection is required from an inferior to a well-disposed superior, others that mere performance of outward duty is sufficient, especially given the dangers of inappropriate obedience to ill-disposed superiors. Common sense, then, remains unclear on whether we have any duty to *cultivate* obedience, even to a good superior.

Sidgwick now turns to his second category: voluntary duties of affection (254–9). The most important, as he sees it, is marriage, any duty to enter into which is derivative from the general duties of prudence and benevolence. Sidgwick notes that the contemporary common-sense view of marriage is that it should be (1) monogamic, (2) designed for permanence, and (3) between non-relatives. At this point he switches away from revealing disagreement to questioning the self-evidence of the relevant common-sense principles. Sidgwick's view is radical, and we also see the beginnings of his argument to the conclusion that common sense is 'unconsciously utilitarian': 'Even against incest we seem to have rather an intense sentiment than a clear intuition; and it is generally recognised that the prohibition of all but monogamic unions can only be rationally maintained on utilitarian grounds' (255). Even as regards permanence, we have to accept that love can decay in a marriage, and it is not self-evident that loveless marriages should not be dissolved. There is also disagreement about whether a marriage requires affection between the parties, and if so how much, and about the duties of each party to the other.

In the case of friendship, Sidgwick meets a new problem: those who deny that friendship should be seen as a matter of duty at all because of the importance of spontaneity. Here Sidgwick could have introduced his distinction between the criterion of rightness and motive to explain how, though friendship does involve duties, friends themselves should not reflect upon them.¹⁵ In fact, he just states the opposing view, that common sense does recognize duties of friendship, the evidence for which is that we often criticize one friend for wronging another.

One thing about the duty of friends to one another is clear: they are more demanding than the duty towards humanity at large. Sidgwick suggests that the only problem here is that arising out of unclarity about what is meant

15. That distinction is indeed implicit in his claim at 242 that duties of special relations will, within utilitarianism, be secondary to the main principle.

by ‘friendship’. It is surprising that he does not point to the obvious fact that common sense does not provide clear and self-evident principles governing the extent to which one friend is required to sacrifice herself to another. Rather, he turns to the questions whether we should try to avoid changes in our affection for friends, and whether such changes affect the duties of that friendship, noting—unsurprisingly—that there is no agreement on these matters.

In the final section of the chapter on benevolence (259–63), Sidgwick turns to his last two categories: gratitude and pity. Sidgwick is happy to accept that there is a duty of requiting benefits. But he then raises several questions to which he claims common-sense morality has no clear answer. Is the feeling of gratitude itself a duty? Often we are inclined to think it is; but we also admire those of an ‘independent’ temperament who find it difficult to bring themselves to feel gratitude. If pressed, we might say that the duty of gratitude requires a willingness to make an equal return. But are we to measure that equality against the effort made by the benefactor (even if the benefit is unacceptable), or the value of the benefit provided (even if the value is, say, one’s own life)?

As far as pity or compassion is concerned, we recognize on the one hand that it facilitates assistance to others, but on the other that it can easily lead us astray into, for example, excessive leniency in punishment. We might also ask whether morality does not require us to give up all our superfluous wealth to the alleviation of the suffering of others. Here Sidgwick shies away from the question at the individual level, arguing that it leads common sense into considering the economic consequences of seeking to provide, either through taxation or voluntary gifts, an income for all, and hence to something at least resembling the utilitarian method (see 436).

Sidgwick’s conclusion on common sense’s capacity to provide clear and self-evident principles of benevolence is, then, pessimistic:

[W]hile we find a number of broad and more or less indefinite rules unhesitatingly laid down by Common Sense in this department of duty, it is difficult or impossible to extract from them, so far as they are commonly accepted, any clear and precise principles for determining the extent of the duty in any case.
(262)

At many points, Sidgwick’s discussion of the duties of special relations is reminiscent of Aristotle’s:

That we should not make the same return to everyone, nor give everything to our fathers, as we do not sacrifice everything to Zeus, is pretty clear. And since we ought to give different things to parents, brothers, companions and benefactors, we should assign to each what is appropriate and fitting... To relatives, fellow-tribesmen and fellow-citizens and the rest we should always try to render what is appropriate, and compare the claims of each in the light of closeness of relation, virtue, and usefulness. The comparison is easier when they are of the same class, and more of a job when they are different; and yet we should not for this reason shrink from the task, but decide the issue as best we can. (1894: 1164a14–35)

Like Aristotle, Sidgwick seeks to clarify the morality of common sense as far as he can. But because that morality cannot provide him with clear and self-evident principles that will provide guidance in every case (see 345–9), he, unlike Aristotle, remains unsatisfied.

His satisfaction comes, of course, only when he has a self-evident principle of rational and impartial benevolence in place, as an underpinning for the common-sense virtue of benevolence. The demandingness of utilitarian impartiality is, in practice, limited by the fact that each person knows his own happiness best and is motivated most effectively by self-interest (431). The pleasures of the natural affections are a vital component of happiness, providing the foundation for more extended altruism, and we are motivated to investigate what will please those we love (433; 439). And because these relationships engender expectations, pain will be caused by disappointing them (439). Entirely impartial ‘enthusiasm of humanity’ is something most of us are incapable of, and we are anyway not in a position to affect more than a very few people (433–4).¹⁶ Indeed, the services to which we are usually prompted are just one example of that whole system of assistance—including the family in particular—required to maintain the normal life of any society (435–7). Gratitude is also supported by utilitarianism, since it helps to promote this system of assistance, and it indeed provides another example of the latent utilitarianism of common sense (437–8). There might seem to be a tension between the view that friends should care for one another disinterestedly and the view that it would be hard to treat someone as a friend if one did not expect gratitude from them for any benefits provided. But utilitarianism recognizes that disinterested beneficence is often required

16. Sidgwick’s complacency on the latter point is unusual for him, but common in the utilitarian tradition. He is quite aware that many see utilitarianism as making excessive demands on the individual: see e.g. 87.

(and is something of which we are often capable) and also that, given self-interest, we would not be adequately motivated to provide benefits in such cases without expectation of something in return. In the end, the only difficulties with benevolence are those, common of course to utilitarianism, of hedonistic calculation (439).

3 Justice

3.1 Justice, Law, and Distribution

Sidgwick's discussion of benevolence consists largely in his pointing out areas of disagreement and unclarity within common-sense morality. The only 'theory' in play is dogmatic intuitionism, the differences between various versions depending on the construction each puts on the specific requirements of the principle of benevolence. The chapter on justice, as we might expect, aims also to suggest points of contention and vagueness. But in addition it engages in significantly deeper discussion of different theories of justice. Sidgwick's discussion foreshadows a good deal of twentieth-century political philosophy concerned with property-based libertarianism and with socialism, as well as stressing certain aspects of justice, such as the importance of respecting natural expectations, which were not picked up to any large extent by later philosophers, including Oakeshottian conservatives.¹⁷ Sidgwick does not discuss Hobbesian contractarian theories of justice as 'mutual advantage' (Barry 1989: 360; index s.v.) in this chapter, although they are in effect covered in the discussion of consent theories in the following chapter (297–8) and, because of their link to freedom, are discussed in the section of 3.11 concerned primarily with this chapter on justice (351).

Sidgwick begins the chapter (264–8) with the suggestion that what he is aiming for, in outlining intuitionist morality, is 'to give to common terms a definite and precise meaning', that is, to provide a definition which all 'competent judges' will accept as 'what they have always meant by the term'. Here, given what Sidgwick says about the interpretation of Bentham (26 n; see 109; CD 1.2.3), we might have hoped Sidgwick would recognize the sense/reference distinction. Presumably, the dogmatic intuitionist is trying to give an account of what justice *is*, rather than to provide a definition of

17. Hayward (1901a: 19; 23) rightly points out the influence on Sidgwick of Mill's chapter on justice in *Utilitarianism*.

the concept. Such a general definition is implicit in what Sidgwick goes on to say as something like ‘the principles governing the distribution of goods and bads’. But the real disagreement between intuitionists will be over what those principles are.

The notion of justice is clearly connected to that of law, Sidgwick goes on, but justice cannot consist in conformity to law. Not all violation of law (e.g. that on gambling) is said to be unjust; actual law itself can often be described as unjust;¹⁸ and there is a domain of justice beyond even ideal law, concerned for example with the relations between parents and children. Which laws are the concern of justice and injustice? Those concerned with the distribution of goods and (in the case of punishment) bads. Equality of distribution is often taken to be just ‘by default’, so it is not surprising to find Sidgwick introducing the link between the two notions. He notes, for example, that it is often thought that a just taxation system¹⁹ will impose equal burdens on all. But, whatever the merits of that view, it cannot be held that a just legal system will affect everyone equally. We allow the law to apportion special privileges (such as legislating, if one is properly appointed) and burdens (such as military service, if one is male). For this reason, some have seen equality in the law merely as a matter of impartial application, but this is to go too far in the other direction: a law requiring all red-haired men to serve in the army, even if enforced with the most scrupulous impartiality, would be held to be unjust. Impartiality matters, but so does the content of law, which should not affect the interests of individuals unequally without good reason. So our question now is what those reasons might be.

3.2 Contract and Custom

So far, Sidgwick has been focusing on justice in the law. He now shifts tack to consider non-legal justice (268–71). First, the distributive nature of justice enables us to see that duties of the affections, in themselves, are not matters of justice; but when they concern the distribution of goods or bads, they are. What counts as justice in distribution in the personal sphere? Just as mere impartiality of application is insufficient to capture the essence of legal justice, so it is not enough to characterize the virtue of justice in

18. Sidgwick points out at 266 that this is one reason why an individual will be interested in the political question of the justice of laws.

19. At 266 n, Sidgwick distinguishes general taxation from payment for specific services.

terms of an attempt proportionally to satisfy all the claims one believes to be reasonable. For reasonableness is an objective matter. Indeed, in the private sphere, some arbitrariness is actually permitted, as in the case of inheritance (268 n).

The most obvious source of reasonable just claims is contractual, to include not only explicit agreements but ‘implied contracts’. As far as the latter is concerned, however, we would not think that anyone is bound by *any* mistaken expectation she knows others will form on the basis of her actions. But if it is one that most would form, then there does appear to be some degree of obligation to fulfil it, though we should note that, because customs vary over time, any obligation resting on custom is to that extent doubtful. Indeed, if the custom is irrational and inexpedient, a failure to respect it may be what is required. In other words, this important area of justice—non-contractual, customary claims—is quite indefinite, which throws into doubt any suggestion that justice is more likely than benevolence to provide clear and self-evident principles governing special relations or other areas of social life.²⁰ Natural expectation conflicts with the notion that a person can be bound only voluntarily and explicitly, and this leads to common sense’s deciding cases purely on the basis of degrees of disappointment:

For instance, if a poor man were to leave one tradesman and deal with another because the first had turned Quaker, we should hardly call it an act of injustice, however unreasonable we might think it: but if a rich country gentleman were to act similarly towards a poor neighbour, many persons would say that it was unjust persecution. (270)²¹

If we return to the law, we find similar unclarity about expectations (271–4). People might be thought entitled to assume that existing law will be maintained, such that when it changes they have some claim for compensation. But both the level and significance of such expectations vary greatly, and extend widely, to the point that not everyone can be compensated, and there is no intuitive principle for deciding between valid and invalid claims. Further, laws can be unjust, and not always as a result of non-legal elements in the social order that might themselves ground conflicting expectations.

20. We should not be surprised by this, of course, given the overlap between the duties of the affections and those of justice already identified by Sidgwick.

21. Sidgwick would probably not have been surprised by the evidence that we appear more inclined to see an action as intentional when its side-effects are bad than when they are good; see Knobe 2003.

Slavery is an example: ‘the conviction that slavery is unjust can hardly be traced to anything in the established order of the slave-owning society’.²²

Sidgwick then launches an attack on the very idea of ‘natural’ expectations, which, he suggests, involves a confusion of the actual and the ideal (see 16–17; CD I.6). On the one hand, ‘natural’ can mean ‘common’ (as opposed to ‘exceptional’) or ‘original’ (as opposed to ‘later’) (see 81). But it can also mean ‘ideal’, in the sense of ‘what would exist in an ideal state’. One line of thought has construed ‘Nature’ in terms of the divine, so that the ideal state is more in line with God’s intentions, and the original state, because he created it, is itself something ideal, any decline from which we humans may be held responsible for. But if we dismiss that view as unhistorical, the tension between the customary and the ideal in our common notion of justice becomes especially clear. Indeed, ‘[i]t is the reconciliation between these two views which is the chief problem of political Justice’ (273). In other words, we are torn in two directions: we want not to disappoint existing customary obligations, but we also want to make our society more just. And these two aims frequently conflict.

3.3 Natural Rights and Freedom

What, then, are the principles of ideal justice? Sidgwick notes that many appeal in this connection to the notion of natural rights (274). Common sense has no clear view on the content of these rights, or how they might be derived. But, Sidgwick continues (274–8), some influential philosophers²³ have maintained that the basic natural right is to freedom from interference, so that the ideally just state of society would involve full recognition of that right and hence ‘equality of freedom’. Note that he has now moved beyond his analysis of common-sense morality to engage directly with non-utilitarian theories of justice.

Sidgwick finds such libertarianism plausible in the abstract, but cannot accept it once he considers its practical application. First, the right must be seen as limited, since presumably we do not wish to rule out interference with children or the insane for their own good. And, since the justification for such interference is benevolent, it is not certain that the same argument

22. Cf. Walzer 1983: 27.

23. Sidgwick probably has Hobbes in mind, calling the view ‘somewhat antiquated’ (see Hobbes 2012: I.14). He says that it is still current, here almost certainly thinking of Spencer (GSM 268). In the twentieth century, the view was most famously developed by Nozick (1974).

cannot be made for interference with the freedom of competent adults. A libertarian will of course just deny this.

If we understand the freedom at stake to be merely that of freedom to act, then the principle seems to allow others to inflict as much pain and annoyance on one as they wish, as long as they do not constrain one's actions. But if we count any such annoyance as interference, it turns out that we cannot prohibit activities causing it without putting unacceptable restrictions on freedom of action itself. The libertarian might here respond to Sidgwick that, since morality extends beyond the fundamental right to freedom, non-action-limiting interferences can be criticized morally, though not prevented if they are not themselves limiting the freedom of action of the victim.

If the right is to be practical, we must assume it protects the right to limit one's own freedom by contract. But if such contracts are to be enforced, it is not clear that this can be justified by reference to *freedom*, since freedom might plausibly be taken to consist partly in an individual's being unable to put any external constraint on her own volitions. Further, if the right is unlimited, it will allow a person to sell herself into slavery, so that the principle is 'suicidal'; and we cannot deduce from the principle any limitations to the right to limit freedom. This is indeed a tricky problem in the theory of freedom (see Mill 1961–91: 18.299–301). But the libertarian might bite the bullet and claim that, since freedom does require the present freedom to sacrifice future freedom, voluntary slavery contracts should be permitted.

Difficulties with the position are compounded when we consider the acquisition of property. It is often thought that the right to freedom includes the right of appropriation, but Sidgwick acutely notes that the most that the right could plausibly protect would be the actual use of something at a certain time: 'the right to prevent others from using at any future time anything that an individual has once seized seems an interference with the free action of others beyond what is needed to secure the freedom, strictly speaking, of the appropriator' (276).²⁴ The so-called 'Lockean proviso' (Nozick 1974: 178–82), that the appropriator leave enough and as good for others, is insufficient to protect the freedom of others, since they may want the very thing appropriated. And the proviso anyway, especially in connection with land, raises the difficult question of the original extent of rights of appropriation.

24. At 277, Sidgwick notes that if the right is to freedom of action, then it will not justify any appropriation.

If we say that someone has a right only to the land he can use, then we have to make precise what we mean by 'use'. Presumably a person's right to freedom does not permit him to exclude others from all land over which she wishes to hunt. And does a farmer have the right to prevent a miner from excavating beneath her fields?²⁵

As I suggested above, the libertarian has answers to Sidgwick's initial problems for her position. But her theory of acquisition does seem to be seriously damaged by Sidgwick's broadside. Indeed, though many of the problems Sidgwick raises for Spencerian theories of the acquisition were addressed by later writers, no plausible account has yet been offered of how, even if we allow a right to freedom as self-ownership, that right can be extended to cover property rights over the world (see Cohen 1995: chs 3–4).

Sidgwick goes on to mention the claims that the right to freedom includes the right to the secure gratification of desires (that is, presumably, the idea that the freedom in question is the freedom to satisfy those desires), and that this right is best protected by appropriation (so no fundamental theory of appropriation is required). Sidgwick notes first that, in a society like ours in which most things have already been appropriated, such freedom cannot be equally distributed. Nor can it plausibly be said that people born into such a society are always better off within it than if there had been no appropriation.²⁶ First, they are not; and, second, even if they are, their freedom has been interfered with, and if compensation is thought adequate, then this suggests that freedom is not the only end of distributive justice.

3.4 Desert

In the next three sections of the chapter on justice (278–94), Sidgwick discusses desert. He begins (279–83) by pointing out that, even if an equal distribution of freedom were possible, we usually understand justice to require the non-arbitrary distribution of all goods, not just freedom. And it may well be that there are good reasons anyway for deviating from equality.

25. Sidgwick also raises a puzzle about the right to bequeath property, which, he suggests, cannot plausibly be said to rest on the right to freedom (since any 'interference' will take place after the person is dead). It could be claimed, however, that actions can extend beyond death.

26. Sidgwick here fails to recognize the so-called 'non-identity' problem (Parfit 1984: ch. 16). But his point could easily be restated in terms of a contrast between the status quo and the abolition of property rights on the person's birth.

To find an overarching principle of justice, Sidgwick begins from the notion of gratitude discussed in 3.4. If we ‘universalize’ gratitude, we obtain the more general principle that people ought to be rewarded according to their desert. For example, if some group of people sets up a profit-making enterprise without any previous agreement on how the profits are to be distributed, it is natural to think that the profits should be distributed according to how much each deserves on the basis of the value of their contribution to the enterprise.²⁷ This principle extends not only to goods, but to ‘bads’, and Sidgwick with some reluctance accepts that it underlies the popular conception of criminal justice as retributive (280–1).²⁸

Some have seen other principles as derivative from a principle of desert, including that version of the ‘individualistic’ principle of freedom discussed in the previous section, according to which the law ought to maximize individual freedom. Here the (rather implausible) thought is that freedom will best enable people to satisfy their desires and so receive what they deserve. Sidgwick suggests that desert also underlies the principle that each is entitled to the product of her own labour. He takes the metaphysically reductive position that labour never really *produces* any material thing, but adds to its value, and notes that merely adding value through labour is not sufficient to ground a property right in some object if that object is already owned by another. All that can be claimed here is *compensation* for the labour, and this, Sidgwick concludes, is what the entitlement principle must mean. Sidgwick’s conclusion, however, need not depend on his reductionist metaphysics. Even if we allow that the Little Red Hen has indeed produced the corn through her own labour, it is plausible to see any entitlement she has to that corn as resting on desert, itself based on the very fact of her labour.

Sidgwick then returns to the Lockean theory of acquisition of property and its use of an analogy between labour and discovery to justify, in particular, ownership of previously unowned land by those who find it. He again points out that common sense will place limits on such rights, based on

27. Whether this is to be measured in terms of effort or results will be discussed at 283–5.

28. Sidgwick is careful to distinguish reparation from retribution, noting (282 n) that in an ‘earlier stage of moral development’ retribution was seen as the standard form of reparation. Sidgwick sees reparation as based (at least in part) on benevolence, which explains the common-sense view that one is under a duty to compensate any serious injury to another, even if one is non-culpable for that injury (281–2; see 253). Since impartial or rational benevolence (that is, utilitarianism) can also justify a retributive system of justice, however, retribution is in the same category as reparation.

what might be considered appropriate compensation for the 'labour' of the discoverer. Finally, he notes that we tend to think that any plan that God might have for the universe will involve the distribution of happiness according to desert, and that human justice should model divine as far as possible.

Because Sidgwick appears to understand desert as based on the value of services, he distinguishes it from another notion, which he calls 'fitness', according to which there are grounds of justice in distribution independent of desert. For example, we think it reasonable that flutes should go to the best flute players, and that jobs should be awarded on the basis of talent. Sidgwick could here have drawn attention to a further confusion within common-sense morality concerning the grounds of desert. In fact, he suggests that the idea of fitness might be seen as a utilitarian principle, rather than as a source of an independent principle of justice.

Sidgwick then moves on to consider 'good desert' (which will always matter, even if crime and punishment disappear) (283–90). The first question is that which arose in connection with gratitude—whether reward should be proportionate to effort or to results (see 261). Here Sidgwick pre-emptively addresses several of the arguments raised in the debates about 'luck-egalitarianism' that developed in the late twentieth century:

For it may be said that the actual utility of any service must depend much upon favourable circumstances and fortunate accidents, not due to any desert of the agent: or again, may be due to powers and skills which were connate, or have been developed by favourable conditions of life, or by good education, and why should we reward him for these? (for the last-mentioned we ought rather to reward those who have educated him).
(283–4)

And, in the case of divine justice, we tend to think that reward should follow only *moral* excellence. But even here the problem of luck enters, since moral excellence itself is at least in part the result of nature and nurture. And once we recognize that, the case for not rewarding non-moral, equally accidental excellences, such as some skill, appears significantly weakened. Indeed, even a capacity for voluntary effort will be said by some to be externally caused (e.g. Rawls 1999: 87), so that the only just outcome will be the default one of equality. Here Sidgwick states the principle at the heart of luck-egalitarianism—that it is unjust that any individual should do better than another merely as a result of circumstances beyond her control—and notes that there appears no reason to restrict the application of this principle to

human beings, which brings ideal justice ‘to such a precipice of paradox that Common Sense is likely to abandon it’ (284).²⁹ Sidgwick is anyway unperturbed by these implications of determinism, since he believes it is practically impossible to separate out those results of a person’s actions which are due to her free choice from those arising from luck (see 72; CD 2.5). So even though we might take reward of effort to be the ideal of justice, in practice we have to content ourselves with rewarding people in accordance with the value of the services they intentionally provide.

Sidgwick notes that we often do put a value on services, speaking for example of what counts as a ‘fair price’. Some may claim that fairness or ‘equity’ in this sense is distinct from justice, but Sidgwick’s view is that, even if there is such a distinct sense, there is another sense in which equity is itself a matter of justice. Often, what is seen as a fair price depends on custom (see 269–70), but this ignores the fact of variations in market value. So can we use market value to measure the value of services, as the principle of freedom discussed in 274–8 would recommend? One problem is that those participating in markets may be poor judges of genuine value. This may be generally so, or true in some particular case. Imagine that I sell you some land in ignorance of the fact that you, as a geologist, have discovered there is probably a valuable mine beneath it. I receive the market value of the land, but not what the land is really worth. Another difficulty is that some goods, such as certain scientific discoveries, though recognized as very important, cannot be priced in a market. But even if we restrict ourselves to cases in which marketable goods are traded under full disclosure, problems arise. Imagine that some monopoly increases the price of its aggregate of services by decreasing the level of available services themselves. It would seem absurd to claim that those rendering the services are consequently more deserving. Or consider a case in which an emergency creates a short-term monopoly: would it be fair for me to offer to save the drowning Croesus’s life at the price of half his wealth? If not, is it ever fair for one party in a market to take advantage of the relatively poor economic situation of another? Indeed, there seems to be no correlation between the market price of one’s services (which may depend on the number of others offering those services, on one’s own degree of willingness to provide them, or on the wealth of the buyers) and one’s desert-level.

29. At 284–5 n. 2, Sidgwick also anticipates, formally, the so-called ‘expensive tastes’ objection to the principle that welfare or happiness should be distributed equally (Dworkin 1981: 228–40).

Given the vagaries of the market, some have proposed a ‘socialistic ideal’, in which levels of compensation for labour would be determined by some set of ‘competent judges’. But what principle will *they* use to assess value? If they value services in accordance with productivity of happiness, they will run into the various problems with hedonism outlined in book 2, as well as further difficulties in comparing the value of necessities with that of luxuries, and estimating the value of each individual’s labour in the cooperative production of any good.

In the light of all these difficulties, Sidgwick concludes, we should give up trying to work out the ‘intrinsic worth’ of services, and value them in terms of the benefits they produce set against the cost of their production (i.e. the reward or compensation offered to their producers). Indeed, we should conclude in general that rewarding good desert is not a practicable goal, a conclusion with which common sense appears to be in large agreement in its condemnation as utopian of any general attempt to distribute rewards in accordance with desert.

What about bad or ill desert, which we have already seen Sidgwick admitting reluctance to accept as part of the morality of common sense? He begins his discussion, in 290–4, by noting that bad desert faces problems similar to those he identified in the case of good. On the one hand, we think it right that a person should receive that punishment required by the law, neither more nor less; on the other, we admit that laws are unjust. There is an interesting disanalogy with civil justice, however, in that, though we might allow that custom may give a person certain rights to unequal privileges, we will not accept that a custom can ground a right to over-lenient punishment.

When it comes to proportioning punishment to the crime, we meet the same problems of causal determination faced by good desert.³⁰ Our practices seem to rest on the assumption of freedom of choice, to the extent that we punish deliberate crimes more than impulsive and we allow upbringing as a mitigating factor. Further, from the moral point of view, the seriousness of some crime seems reduced if it is performed from a worthy motive, though common sense will hold that motivation should not be taken into account in judicial practice. But even if we focus just on intention, rather than motive, it is difficult to state principles for judging the seriousness of a crime. Sometimes a criminal’s intention is to do good, or at least not to

30. Again, despite the main claim of 1.5, Sidgwick seems ready to accept the potential practical relevance of the free-will debate here. See CD 2.5.

harm others. We tend to think a crime's being intended to be secret is irrelevant, though we allow that a good deal of the harm done by a crime is the insecurity caused by its being known. And, in an interesting remark which suggests that Sidgwick may have allowed for the possibility of 'rule' or 'generalization' forms of utilitarianism, Sidgwick notes:

[I]f we say that the heinousness of the crime depends on the loss of happiness that would generally be caused by such acts if they were allowed to go unpunished, and that we must suppose the criminal to be aware of this; we seem to be endeavouring to force a utilitarian theory into an intuitional form by means of a legal fiction. (292)

Sidgwick then returns to the issue of negligence (see 60; CD 2.3.1), noting that the criminal law sometimes punishes pure 'carelessness'—that is, a mere failure to act, which is 'at the time strictly involuntary' and so not a matter of ill desert. Here the only justification for punishment, Sidgwick claims, can be the utilitarian one of deterrence. Questions similar to those concerning desert and negligence will arise about how to determine the limits of reparation, if we allow that someone can be entitled to reparation as a result of harm caused by another's negligence.

Sidgwick's complaints about common-sense principles of justice at 349–52 are those one would expect from the discussion in 3.5. Should desert be proportionate to effort or size of benefit? How many of a person's qualities are due to themselves, and how much to their nature and education? How should we decide conflicts between different principles of justice, such as those between fitness and desert, or between these principles and the libertarian principle? That latter principle is itself deeply problematic. It cannot justify appropriation or the enforcement of contracts, nor can it be saved through reliance on a fictitious social contract. Further, common sense is unclear on how much weight to place on customary expectations in any move towards ideal justice.

4 Laws and Promises

4.1 Political Obligation

Sidgwick begins his chapter on laws and promises by repeating his earlier point that, though the notions of justice and law are closely related, they do not overlap completely (295–7; see 265). Further, because injustice involves

the violation of certain rights, and some failures to keep promises are often thought wrong independently of any injury to another (such as promises to the dead), promising also deserves a separate discussion. This seems to be another place in which Sidgwick allows his own philosophical views to skew his interpretation of common sense, since many would claim that breaking a promise to a dead person is somehow a wrong to that person. But nothing of consequence hangs on this: one can see 3.6.5–9 as a continuation of 3.5 if one wishes.

Sidgwick begins with a discussion of political obligation—that is, the moral obligation to obey the law—understanding the law in question to consist in ‘Rules of Conduct laid down by a Rightful Authority, commanding within the limits of its authority’ (296). This raises two related questions: (1) How do we decide who or what is the rightful lawmaker? (2) What are the limits of the lawmaker’s authority? To the first question, Sidgwick suggests three possible answers: the ideally right government; the *de iure* government (that in accordance with the constitutional traditions of the relevant country); or the *de facto* government (a government which is established but not in accordance with constitutional tradition).³¹ The first to be discussed is the ideal (297–9).

Sidgwick begins with the consent theory of political obligation, according to which individuals are bound to the law through some kind of social contract (see 274–8). Any such contract must be to some degree ‘tacit’ (see Locke 1988: 2.119–20); it might be claimed, then, that by remaining in a community any person is undertaking to obey the laws of that community.³² This, Sidgwick suggests, is to allow the ideal to collapse into the customary, in the sense that *any* established government, however despotic, could demand obedience. He does allow that a consent theorist might incorporate natural rights into her theory, so that any government violating these rights will be illegitimate. But we then face the problem of finding some self-evident principle from which these rights can be deduced. No such principle can be founded on freedom, as we have seen, and even if it could, it would not support the right to private property.

31. On the distinction between the ideal and the customary, see 299–301; also 16–17 (CD 1.6); 272–3 (CD 6.3.2).

32. Sidgwick does not mention any explicit commitments entered into by immigrants into a community, presumably because he is seeking a comprehensively binding theory of political obligation. See 303.

What about actual consent? Rousseau's notion of the 'general will' is not up to the task of justifying political obligation: '[I]t is paradoxical to affirm that the freedom and natural rights of a dissentient minority are effectively protected by establishing the condition that the oppressors must exceed the oppressed in number' (298). Further, any principle of actual consent, if it is to be 'absolute', must apply to everyone. But if we exclude children, we are drawing an arbitrary line, and will have to extend the suffrage to women.³³ Nor can representative democracy plausibly be claimed to realize the ideal of 'obeying oneself', since usually a representative assembly is chosen only by a subset of the citizenry and any law is approved only by a subset of the members of that assembly. Finally, understanding the ideal itself in terms of majoritarianism goes against the Socratic view that laws should be made by experts, and this latter principle has as much claim to self-evidence as the former.

Sidgwick then moves on to the question of when it is acceptable to introduce an ideal constitution in violation of the established order in any community, and the suggested non-ideal sources of authority (299–301). He notes that there is no agreement on the first question, and then points out how difficult it can be to say what the established government actually is (after a coup, for example) and when it should be obeyed in general. And even in a constitutionally governed state, there is disagreement about whether subjects have the right to disobey unconstitutional commands or even in such cases to prosecute the sovereign for violating the constitution. It is very hard to work out what constitutional obligations really are because the history itself is unclear, because certain restrictions on sovereignty might have been the results of coercion and hence not binding, and so on. What tends to happen is that people's views of the nature of their actual constitution are influenced by their views of the ideal.

33. Sidgwick himself (EP 379; 384–7) doubts the utilitarian argument that the interests of women in general can be represented by their male relatives, and the arguments based on women's physical inferiority to men that (1) the right to vote belongs only to those who will defend their country and (2) extending the vote to women creates a danger that a physically more powerful minority may refuse to abide by the decision of the larger majority. He thinks the arguments for extending the franchise to wives 'appear less strong and the objections more serious'. The interests of wives will usually be looked after by their husbands, a wife is likely to vote along with her husband for the sake of domestic harmony, and the usual division of labour gives husbands a greater range of experience on which to base their vote as well as making it impossible for many of them—e.g. those in the military—to use their vote, so that women would have disproportional electoral power. He thinks also that the appeal to the possibility of disobedience based on physical power only becomes 'serious in the extension of the franchise to wives'.

Sidgwick now focuses in more detail on the issue of the limits to obedience (301–3). The suggestion that immoral commands should be disobeyed is unhelpful because tautological. The question is how to decide which commands are immoral, and here there is no clear principle to guide us. For example, people will differ over whether a son is permitted to assist a parent fleeing from punishment, or whether certain unjust commands by some government ought to be obeyed. The only way to decide these and similar issues is through the application of utilitarianism.³⁴

4.2 Promises

Sidgwick's discussion of promises can plausibly be said to be the best ever, and is a highlight of book 3. He begins by noting that, though the social contract is a fiction, it is nevertheless the case that a good deal of constitutional law has been based on agreements between different sections of the community and the authority of a lawgiver has sometimes been increased by explicit oaths of allegiance (303–4). So this discussion is linked to that of political obligation in the first half of 3.6.

Some philosophers, Sidgwick points out, have seen the duty to keep promises as identical to the duty of veracity.³⁵ He accepts that there is an apparent analogy between them, since both have to do with the relation between facts and statements: promising involves making facts fit one's statements, and veracity making one's statements fit the facts. But even this analogy is weak, since we are not required to make the facts fit all our statements, such as those merely announcing an intention. What matters with promising is not conformity to what I have said, but to the expectations I have intentionally created in other people. What if people understand my promise in some way I did not intend? Here, common sense is likely to see me as bound by what I have said if the interpretation people have put on my words (or other relevant signs) is a natural one. But this is a matter of justice in general, not fidelity to a promise, since no promise has been made. In fact, this claim of Sidgwick's is dubious. The meaning of a sentence does

34. This is an excellent example of Sidgwick's use of the kind of argument Mill made against non-utilitarian accounts of justice (1998: 5.26–31): the only way to avoid disagreement on the limits of obedience is to adopt utilitarianism. But of course this argument can be offered by the proponent of any position in the debate; see Donagan 1992: 132. Utilitarianism may be said to have the advantage that it can explain apparently competing principles as themselves grounded in utilitarianism; but others may have their own explanations.

35. He has Wollaston, and perhaps Clarke, in mind (OHE 198).

not depend on the speaker's intentions, and indeed Sidgwick goes on to weaken his claim, allowing that we may say that there has been a promise, albeit an 'imperfect' one.

Sidgwick allows that the principle that perfect promises should be kept seems certain to common sense, prior to reflection. But he then goes on to point out many ways in which the principle must be qualified:

1. Since a promisee can cancel a promise, common sense needs to take a view on cases in which the promisee is dead or otherwise incapable of granting release. (305)
2. It must be recognized that promises to perform immoral acts are not binding. (305)
3. The extent to which promises override prior obligations requires clarification. For example, is providing promised financial aid to a friend morally more important than giving one's children a good education? (305)
4. Clarification is required on the bindingness of promises obtained by force or deception, and those based on misunderstanding. (305–6)
5. It must be explained how far the foreseen effects of keeping the promise must change for the promiser no longer to be bound. (306–7)
6. Clarity is required concerning cases in which the promisee is inaccessible and the fulfilment of the promise is against the wishes of both parties. There is much dispute in particular about whether we are required to do what *would* have been the promiser's intention. (307)
7. It must be decided whether a promise can make it right to harm anyone, and, if so, when. For example, what if the fulfilment of the promise will harm a foolish promiser herself? Or if the promisee will be harmed? (308)
8. It must be explained what are the duties of a promiser who unintentionally creates certain expectations, when what is typically and customarily understood is itself unclear. (308–9)
9. The bindingness of promises made to the community, when they are not interpreted in the same way by all, must be clarified. (309–10)

Sidgwick concludes his discussion as follows:

[I]t appears that a clear *consensus* can only be claimed for the principle that a promise, express or tacit, is binding, if a number of conditions are fulfilled: viz.

if the promiser has a clear belief as to the sense in which it was understood by the promisee, and if the latter is still in a position to grant release from it, but unwilling to do so, if it was not obtained by force or fraud, if it does not conflict with definite prior obligations, if we do not believe that its fulfilment will be harmful to the promisee, or will inflict a disproportionate sacrifice on the promiser, and if circumstances have not materially changed since it was made. If any of these conditions fails, the *consensus* seems to become evanescent, and the common moral perceptions of thoughtful persons fall into obscurity and disagreement. (310–11)

This is a crucial passage for the defender of dogmatic intuitionism. Sidgwick later admits that ‘the duty of fulfilling express promises and distinct understandings’ may ‘properly claim universal acceptance as self-evident’, if the promises in question are both understood in the same way by promiser and promisee, the promise can be cancelled by the promisee, and is not in conflict with prior obligations (352–3). But, he goes on (353–4), this is not the case with the other qualifications, concerning, for example, promises based on false statements. Here there is no clear consensus, and the qualifications point in the direction of utilitarianism. But the dogmatic intuitionist can restrict her principle to cases in which none of these qualifications are relevant, and continue to claim ‘universal acceptance as self-evident’.

Sidgwick has one final argument against such a move (354). When we reflect upon the ground of the duty to promise, we will see it rests upon the potential harm of disappointed expectations, not upon the very fact of having promised. Now, the dogmatic intuitionist may indeed agree that one of the reasons why breaking a promise is wrong is that it risks disappointing expectations. But her intuition will be that such infidelity is wrong in itself, independently of the harm of disappointment (see Phillips 2011: 105–6).

4.3 Utilitarianism and Justice

At 439–48, Sidgwick covers the topics raised in both 3.5 and 3.6 from the utilitarian perspective. The utilitarian value of ‘Humean’ justice as ‘order’ in society is obvious, and the limits of the duty to obey the law are to be determined by utilitarianism (440). Indeed, utilitarianism explains the disagreements arising between conservatives and reformers, since it recognizes both the value of a reasonably effective *de facto* government as well as the potential for improvement (441; 447–8). Utilitarianism, though it recognizes universalizability, explains why certain inequalities between individuals are

not unjust (442–3). Freedom is a source of happiness to agents, as well as grounding motivation, and individuals should often be left to distribute benefits according to their own personal preferences. Utilitarianism not only recognizes the significance of meeting normal expectations, but, unlike common sense, need not specify whether such expectations must be, say, based on definite agreements (442). Further, utilitarianism can explain the various exceptions and qualifications to principles of fidelity (442–4). For example, if circumstances change after a promise has been made, the expectations are not those originally created and the utilitarian reason for keeping the promise is that much weaker. Utilitarianism can also identify the advantages in each of two competing conceptions of distributive justice: the individualist conception, grounded on freedom, and the socialistic conception (444–6). Freedom enables each person to seek the happiness of the individual she knows best and is most concerned for (i.e. herself), but there must be limits to such freedom and utilitarianism provides them. Further, the limits to a free market grounded on desert, accepted by common sense in cases in which, for example, individuals take advantage of their superior knowledge, can be justified by utilitarianism. Common sense as well reflects the utilitarian view that the rewards of ‘good desert’ should be proportioned not only to the value of any benefit provided, but also to the difficulty of motivating people to provide them (446–7). And if there are no such considerations in play, utilitarianism supports the principle of equality as defeasible, as does common sense (447). In general, then, utilitarianism provides a way to resolve the conflicts between different principles of justice (447–8).

5 Veracity

It might be thought that ‘social duty’ as a whole is covered under the headings Sidgwick has already employed: justice, benevolence, and law and contract (312–13). Whatever is *owed* to others (except as matters of law and contract) could be said (if we extend common usage slightly) to be a matter of justice, and all other more indefinite obligations to be matters of benevolence. We might then go on to distinguish self-regarding duties from these social duties.

Sidgwick first objects to the very distinction between social and self-regarding duties, on the ground that they imply that the end of right action

is always happiness, either that of others or that of the agent. This would leave no room for the key intuitionist idea that certain kinds of conduct are required in themselves, without reference to the happiness or unhappiness in their consequences. It is not clear why Sidgwick thinks that the social/self-regarding distinction is naturally interpreted hedonistically. But he anyway allows that one might include non-hedonistic effects on others or on the agent, objecting that these also play no clear role in the common-sense understanding of certain moral rules. Further, he suggests, our actions often affect both ourselves and others, and it is difficult to decide which effect is more important. Why this is a problem for the distinction, however, is not clear, since presumably some action could be said to violate both self-regarding and social duty. More problematic is the point Sidgwick goes on to make—that certain virtues, such as courage (see 332), cannot easily be slotted under either heading.

The original claim, however, might be revived at this point. Perhaps the virtues Sidgwick has already discussed cover everything in morality except the territory of the purely self-regarding virtue of prudence. Sidgwick has already noted that common sense appears to require certain actions, understood entirely independently of their consequences, and he must be understood to be including non-self-regarding actions in this category. But it can be claimed that veracity, though its moral force does not rest entirely, or indeed perhaps at all, on the effects it has on others, is nevertheless a matter of what we owe to others. So Sidgwick appears not to have provided a watertight response to this revised version of the original claim. Veracity could be seen as matter of justice, broadly construed.

The more important issue here, of course, is the ground for the duty of veracity itself. Sidgwick claims (313–14) that he has already shown that many common-sense moral principles rest only on the promotion of welfare, but that there are important apparent exceptions, one of the most important being veracity.

He notes that the analogy between veracity and the duty of fidelity to promises—that both require a correspondence between words and facts—makes it appropriate to discuss veracity at this point, since similar questions arise in each case (see 303–4; CD 6.4.2). So, just as the duty of fidelity requires one to keep promises as one knows they have been understood by the promisee, so too veracity requires us to use words which we believe will produce in others beliefs similar to our own. This seems to be a helpful clarification of the common-sense duty, but problems again arise from the

use of set forms based on law or custom (see 309–10). Should we understand, say, declarations of religious belief as they were originally taken, or as they are now commonly interpreted?³⁶ Most would assume that phrases which are never now literally interpreted (such as ‘your obedient servant’) are not objectionable, but what of cases in which people might still be deceived (as for instance when an inconvenient visitor is told that one is ‘not at home’)?

These problems, however, are limited. In most cases, it is quite easy to see what is required by the principle that one should speak the truth. Nevertheless, that principle does not provide an indubitable example of an ethical axiom (315–17). The first problem is that it is not clear whether the principle of veracity is derivative from some more fundamental principle, such as a principle of dignity, or a principle of rights.³⁷ If the latter, then it would seem that lying in one’s own defence is likely to be approved of by common sense, which will also permit the use of falsehood by legal advocates, as well as ‘white lies’ to invalids who might be harmed by the truth or, concerning certain matters, to children. Revealingly, Sidgwick says that he sees no way of deciding when deception is justified ‘except by considerations of expediency; that is, by weighing the gain of any particular deception against the imperilment of mutual confidence involved in all violation of truth’ (316). He will not allow that the very fact that one is lying, or that a person is being deceived, might itself be put into the balance as a morally relevant consideration.

Sidgwick then raises the issue whether common sense will allow the use of falsehood in religion with a view to generating true belief (316–17). He suggests that some will indeed permit such falsehood, though the standard view is that veracity requires one only to make one’s actual statements true. Sometimes concealing the truth will be called for, however, and here, though one must not lie, one may indirectly produce a false belief through a misleading answer. Some, however, fail to see any morally relevant difference between this and direct lying.

In other words, though it is usually clear what veracity requires, there are enough cases in which it is not clear to rule it out as a moral axiom (see 354–5).

36. After the failure of a movement to reform the University of Cambridge, and in particular to bring it about that fellows on their election no longer be required to promise to conform to the 39 Articles of the Church of England, Sidgwick resigned his fellowship in 1869. See Sidgwick and Sidgwick 1906: 172; 196–202; Schultz 2004: 124.

37. Sidgwick here quotes Kant 1871: I.I.2/267.

At 317–19, Sidgwick turns to an argument, Kantian in inspiration (Kant 2002: 1.14 [4: 402–3]/203–4), for an absolute principle of veracity:

It is said that if it were once generally understood that lies were justifiable under certain circumstances, it would immediately become quite useless to tell the lies, because no one would believe them; and that the moralist cannot lay down a rule which, if generally accepted, would be suicidal.

The contrast here is between absolute and non-absolute principles of veracity, so the kind of non-absolute rule that Sidgwick has in mind is, presumably, a qualified principle requiring one not to lie, except in certain circumstances. If it then became pointless to lie in those circumstances, the argument continues, no one would do so, and so the rule would never be applied.

Sidgwick has three responses to this argument. First, he claims, it is not always bad if trust breaks down; indeed, it may be for the best, as for example in cases of legitimate secrets, where it is understood that those illegitimately seeking access to them will be fobbed off with lies. In such cases it may indeed turn out that lies are no longer told, but this result is not in itself bad and so provides no reason why falsehoods should not be told in the process leading to the situation in which they are never told.

This would be a good response if the original argument were based on the badness of trust's breaking down. But the argument can be understood as methodological: there is something close to self-defeat in recommending a principle which, if generally accepted, would never be applied.

Sidgwick's second response is more powerful: that stating falsehoods can be effective even when it is generally understood to be legitimate, as in the case of legal advocacy. The main reason for this, however, is not that given by Sidgwick—that beliefs are 'not formed purely on rational grounds'. If one knows an advocate to be stating a falsehood, one will not believe her. The reason why we do not disbelieve such advocates in certain cases is that we do not know that they are lying, and the reason we tend to believe them is that even advocates tend to tell the truth. Consider an adaptation of another of Sidgwick's cases: that of the invalid, wishing to know the truth in a world in which everyone accepts the principle that one is permitted to lie to invalids. Here again, any particular invalid will not know whether the person from whom she is seeking information is choosing to exercise her prerogative to lie. So the questions may be asked, and lies may be told. What does seem impracticable is a rule that is known to be accepted by all which

requires lying to invalids. The effect of that would be that invalids would not bother to ask the relevant questions in the first place. Here, however, the non-absolute theorist may still recommend a rule requiring lying *if one is asked*. That rule is always applicable, even in a world in which no one ever asks.

Sidgwick's final point is the standard response to Kantian arguments that we should not assume that, even if the universal application of some maxim would be very bad, it is never right to act on it. For it may be that the maxim is not universally accepted, and that one's acting on it will not tend to bring it significantly closer to being so. Again, however, this seems to assume that the argument is based on the badness of universal acceptance, rather than the incoherence of recommending a rule which would never be applied if generally accepted.

Sidgwick's brief discussion of veracity in 3.II ends with the claim that it is even more clear than in the case of fidelity to promises that the numerous limitations of the principle of veracity are 'very commonly determined by utilitarian reasonings, implicit or explicit' (355). The discussion in book 4 expands a little on this, noting for example how common sense agrees with utilitarianism in allowing deception of a criminal with careful attention to any deleterious effect such deception might have on the disposition to honesty of the deceiver (448).

6 Malevolence and Liberality

In 3.8, Sidgwick turns to 'minor social duties and virtues'.³⁸ The first set of such duties (320–4) could not be placed under the heading of benevolence, in that they concern malevolent feelings, such as envy. One difference between such feelings and those of benevolence is that human beings typically feel some default degree of kindness to others. Nevertheless, there are some structural analogies. Just as we love those who benefit us, so too we dislike those who have harmed us, even if the harm is morally permitted, as in the case of a competition won by a rival.

38. He now appears to accept the social/self-regarding distinction. This may be because he is prepared to allow for a third category of virtues, such as courage and perhaps veracity, which are neither purely social nor purely self-regarding. Courage is perhaps a hybrid, whereas veracity moves beyond the distinction entirely.

On the question of the rightness of such feelings, common sense is unclear. Some say that they are always to be repressed, and indeed most of us would reject all kinds of envy. Further, we admire the virtues of good humour, meekness, mildness, placability, and mercy. But common sense also recognizes the legitimacy of instinctive resentment for wrong. Nor can the matter be left to the courts, since there are times when effective requital for some injury is possible only privately. Christians recommend forgiveness in such circumstances, but even they are likely to accept that some deterrent punishment is justified in certain cases. Once again we find Sidgwick suggesting that the only way to decide such questions is through a broadly utilitarian method, which requires consideration of whether the punishment being considered is really beneficial to society.

And the same goes, he suggests, for malevolent feeling.³⁹ Some hold that anger should always be directed at the act, not the agent, towards whom we should be benevolent; while others believe that this combination of malevolence and benevolence is too complex for it to be an ordinary duty. Some allow for instinctive, but not deliberate, resentment, urging that any action prompted by the latter could have been the result of a more worthy motive; and others believe in the value of a desire to punish wrongdoers as one of the foundations of a system of justice. Further, it may be claimed that we can distinguish that morally legitimate desire from the illegitimate desire of the pleasure to be gained from the infliction of punishment. But if it is important 'for the well-being of society' that people should feel such pleasure at justified punishment, such a desire is bound to motivate, and all we can say on reflection, perhaps, is that people should not gloat over the expected pleasure. Sidgwick's final judgement, anticipated here in 3.8 through the link drawn between reflective common sense and societal well-being, is that the different reflective views taken of malevolence in fact correlate with different predictions about its consequences (449).⁴⁰

Sidgwick then turns to the remaining social virtues, suggesting that they are all derivative from virtues he has already discussed (324–6). He chooses liberality as his example, both because of its historic importance (see e.g. Aristotle 1894: 4.1) and because of an interesting complexity in the feelings

39. Sidgwick is himself clear about his lack of impartiality in 3.8 (remember that book 3 is intended as a neutral survey of common sense, identifying its problems rather than providing any solutions). At 355, he says that in that chapter 'it was made plain . . . that as regards the proper regulation of resentment, Common Sense can only be saved from inconsistency or hopeless vagueness by adopting the "interest of society" as the ultimate standard'.

40. This claim is extended to other virtues at 453–7.

it involves. Sidgwick claims that liberality, as a virtue, is merely benevolence as applied in particular to the giving of money, beyond the limits of strict duty—it is a ‘graceful excess’. Note that Sidgwick is not suggesting that liberality is supererogatory. If I am rich and give only the minimum, I cannot be blamed for lack of benevolence, strictly understood, but I can for lack of liberality. The excess is needed for the duty to be ‘well done’. Sidgwick then suggests that this graceful excess is possible only for the rich. But there seems no reason for this restriction. The widow’s mite may well represent the upper end of what she could spare, and her duty may be well done. His suggestion, however, that there is something aesthetic in our admiration of the ‘careless profusion’ of the rich, and in our disapproval of meanness (324–5), is more plausible.

Sidgwick goes on to note that liberality also operates in the borders between justice and benevolence, in that a liberal person is one who will satisfy customary expectations fully, even when they are indefinite. Liberality comes into play, Sidgwick suggests, only when the stakes are relatively low. A great sacrifice will count as generous rather than liberal (325 n). The same goes for the vice opposed to liberality—meanness.⁴¹ If the stakes are high, it is not mean to claim one’s own rights, and worse than mean to deny another their rights. Meanness, that is to say, is not injustice; the mean person ‘chooses a trifling gain to himself rather than the avoidance of disappointment to others’ (325). It is interesting that, on Sidgwick’s view of common-sense morality, vices such as meanness and injustice do not overlap. He could perhaps have noted here another unclarity, however, since presumably some will want to allow that an action could be both mean and unjust.

Sidgwick concludes his categorization with the suggestion that generosity, in so far as its sphere overlaps with that of liberality, transcends liberality, is more a matter of feeling than outward acts, and involves a greater suppression of selfish impulses. In its wider sense, it is found in competition of all kinds, where it is sometimes called *chivalry*, and consists in the exercise of benevolence when it is especially difficult and hence especially admirable. For example, a chivalrous competitor may surrender a chance advantage. But this is not seen as a strict duty, and ‘what some would praise and approve, others would regard as quixotic and extravagant’ (326).

41. At 326, Sidgwick suggests that meanness is more appropriately seen as the opposite of generosity, since it concerns not just money but the taking of advantages in general.

7 Self-regarding Virtues

Sidgwick begins his discussion of the self-regarding virtues with the—to modern readers somewhat surprising—claim that, since common-sense morality assumes a harmony between self-interest and virtue, the performance of duty in general is equivalent to performing a duty to oneself to promote one's true interest (327).⁴² He then goes on to point out that this assumption of common sense is at odds with its postulation of a duty not to commit suicide, even if one's life contains a preponderance of pain over pleasure. Less surprising is his ascription to common-sense morality of the view that we have a duty of prudence to promote our own happiness, except in so far as by sacrificing it we can promote the happiness of others.⁴³ He then goes on to suggest (327–8) that, since it might seem that people already desire their own good enough (without any need for the promptings of duty), prudence as a virtue is often seen as intellectual, its scope being the calculation of means to ends and the resisting of irrational impulse. As Sidgwick points out, prudence then amounts to wisdom (as he understands it), with self-interest as an exclusive end. Sidgwick here appears to assume—again, implausibly—that common sense will take it to be reason, rather than rational desire, that opposes irrational desire.

Sidgwick then moves on to consider particular virtues which *may* be independent of prudence, the most significant of which is temperance, understood as the habit of controlling one's appetites (328–9). On certain aspects of temperance there is broad agreement: the appetites should be checked if their satisfaction is a threat to bodily or mental health; the satisfaction of an appetite is imprudent if it involves the loss of some greater gratification, and wrong if it interferes with the performance of duty.⁴⁴ On what we might

42. Anthony Skelton suggested to me that Sidgwick probably had in mind the theistic element in common-sense morality, according to which virtue will always be rewarded, and vice punished. This is borne out by 391; GSM 188. See also Whewell 1845: 1.374; Parfit 2011: 1.142.

43. Here, however, was another opportunity for Sidgwick to note an unclarity in common-sense morality. Mill's view in *On Liberty* 1.9 that we should use the power of morality to criticize people only in so far as they are harming others (1961–91: 18.233) clearly struck a chord in the common sense of Sidgwick's time. See the following note.

44. Sidgwick doubts whether this latter kind of case would usually be criticized as 'intemperance'. If he means that we might focus rather on the non-performance of duty (so we might describe the case as one of, say, injustice), he may be right, though I suspect that accusations of weakness and so on are quite standard in such cases. If, however, he means that common sense will not *view* such cases as matters of intemperance, this seems mistaken.

call the *permissive view*, implicit in Sidgwick's text,⁴⁵ appetites can be gratified within these limits. But on this view itself there is no agreement, and two other views are suggested. The first is that appetite should never be gratified beyond the point required for health, while the second lies between this extreme and the permissive view, allowing gratification in so far as it promotes certain other ends, such as sociability.⁴⁶

In 3.9.2, Sidgwick has in mind mainly the appetites for food and drink. He begins the next section with the claim that the regulation of sexual appetites by common sense clearly goes beyond mere prudence. His meaning is unclear until a page later, where at 330–1 he admits that the virtue of purity is not purely self-regarding and that he has been led to discuss it in 3.9 only through its overlap with temperance. He is not prepared, however, to accept the claim that purity, since it concerns propagation, is entirely social. In this respect, then, purity is like veracity: a virtue understood by common sense as at least partly independent of any promotion of the happiness of the agent or others.

Again, we find a good deal of disagreement about purity within common-sense morality (330). One immediate view might be that the bounds of purity are set by the law. But there can be legally permitted impurity, and some illegal sexual intercourse is not thought to be impure (e.g. that in a marriage in which the parties had deliberately omitted to fulfil certain legal conditions: 357). No one accepts a view of purity analogous to the permissive view of temperance. But two different views mirror the latter two views on temperance: a stricter view allows sex purely for propagation, while a less strict view permits it also as promoting mutual affection. Sidgwick also notes here an interesting example of potential self-effacingness within common-sense morality itself (see CD 3.3.4; 7.2.3):

[A]ny attempt to lay down minute and detailed rules on this subject seems to be condemned by Common Sense as tending to defeat the end of purity; as such minuteness of moral legislation invites men in general to exercise their thoughts on this subject to an extent which is practically dangerous.

Sidgwick's own open-minded attitude towards the Victorian position on sexual morality emerges most clearly, and somewhat impatiently, at 359.

45. Note that the position outlined at 329 is described as a 'third...view'.

46. We should assume, given what Sidgwick goes on to say about the attitude of holders of the intermediate view to 'solitary...pleasures of the table', that they will not allow the agent's own pleasure per se as a legitimate end.

On reflection, we will see that this area of morality is concerned with the orderly continuation of humanity, and the protection of dispositions of feeling with a view to perfecting individuals or making them happy (see 450–3). These aims can conflict, and even when they do not it is certainly not self-evident that ‘repression of sexual license’ is a necessary means to either.

When it comes to desires for most things apart from food, drink, and sex, Sidgwick suggests, no special issues arise (331). There is a general duty of self-control, and common sense does not, in cases involving, say, the enjoyment of bathing, accept the ascetic position that any sensual enjoyment is bad. The only exception is the desire to avoid pain, and here (as in the case of its prohibition of suicide) common sense sees it as admirable to endure pain beyond the level most conducive to the happiness of the agent or others. And this brings us to courage and fortitude.

In the case of the self-regarding virtues, Sidgwick is less convinced that utilitarianism is working behind the scenes in common sense (356; 449–50). Self-control may be seen as subordinate to benevolence, but also to prudence, some other rule of social duty, or some broadly moral end. But despite the unclarity, there appears to be ‘no obstacle’ to our defining the end here as happiness, even in the case of extreme asceticism. It may seem that purity cannot be consistent with utilitarianism, given the restrictions it places on pleasure-seeking. But it turns out that ‘it shows a specially complex and delicate correspondence between moral sentiments and social utilities’ (450).

8 Courage and Humility

Courage, Sidgwick says, we tend to see as the disposition to face not danger in itself, but the danger that arises in the performance of duty (332–3; cf. 219; CD 191).⁴⁷ And in working out what our duty actually is, we have to take into account the pain and danger to ourselves. So it will be generally agreed, for example, that we are not required to endure any pain except for the prevention of much greater pain to another or for some much greater good.⁴⁸ But we have already noted the disagreement about the demands of

47. Sidgwick distinguishes fortitude as concerned with the endurance of pain, though allows that this is also sometimes seen as the domain of courage also.

48. Common sense will presumably also require me to endure some pain so as to prevent many people experiencing a slightly less serious pain.

benevolence,⁴⁹ and that disagreement will of course carry across to our analysis of courage (see 355).

Sidgwick then (333–4) considers courage as an excellence, rather than a matter of duty, where its instantiation is often not voluntary: danger often arises suddenly, so one's response is not a matter of willing; and even when it is, one may be too timid even to be able to exemplify the admirable impetuosity and vigour of the truly courageous person. Sidgwick is in no doubt that we often see such behaviour as morally admirable in itself, but only up to the point at which it seems rash. How are we to locate that point? The only reflective principle Sidgwick can find within common sense is the utilitarian one:

On utilitarian principles we should endeavour to strike as exact a balance as possible between the amount of danger incurred in any case and the probable benefit of cultivating and developing by practice a habit so frequently necessary for the due performance of important duties. (334)

Sidgwick does not consider a view according to which, though the limit depends to some extent on the consequences of courage, the exercise of courage has some value in itself.

Sidgwick closes his survey of common-sense morality with a discussion of the Christian virtue of humility, including it alongside courage as another virtue which is praised independently of its effects on happiness (332; 334–6). Sidgwick sees the common conception of humility as paradoxical, since it requires us to have a low opinion of ourselves even if our merits are high. He suggests another way of understanding humility, which requires not ignorance of our merits but a readiness to compare ourselves with our superiors, or an ideal, rather than our inferiors. But, he claims, this suggestion ignores the fact that, in central choices concerning how we are to spend our lives, we must compare ourselves carefully with the 'average'.

Sidgwick does not explicitly consider the view that we should be disposed to compare ourselves with superiors *except* when making important life choices, but goes on to suggest an interpretation of humility close to it, which requires us to repress self-admiration. This principle, however, is not seen as ultimate, since it is usually justified on the ground not that self-admiration is bad in itself, but that it impedes the development of virtue. Further, there is potential tension between this conception of humility and the view that self-respect—especially moral self-respect—is important. Self-approbation,

49. At 333 n. 2, Sidgwick refers back to 3.4.5, probably having 252–3 in mind.

that is to say, should be seen as permissible up to the point that it begins to impede the development of virtue.

The same tension arises in the case of the respect of others, the desire for which is seen as valuable to some extent as an aid to moral development. We are required to repress that desire on occasion, but this is not a matter of humility. What humility requires is that we control any desire we might have to claim such respect from others. But even here there is a limit, since sometimes the failure to offer the marks of respect is an insult to which dignity, for example, might require us to respond. And there is no agreed principle to determine those limits.

9 The Failure of Common-sense Morality

In 3.11, after his long survey of common-sense morality in 3.3–10, Sidgwick undertakes to bring together the various threads of his argument. He begins by reminding us of his general aim (337–8): to decide whether common sense can provide the basis for ‘clear and precise principles commanding universal acceptance’, on the understanding that such principles may be arrived at only through philosophical refinement of common-sense morality itself. These principles, in so far as they are self-evident, must of course be non-derivative and ultimate (350–1).

At 338–42, Sidgwick sets out the four conditions the complete fulfilment of which by any apparently self-evident proposition will make that proposition as certain as it can be (see CD 4.3.1): (1) clarity; (2) reflection; (3) consistency; and (4) non-dissensus. As we have seen, in the earlier chapters of book 3, Sidgwick has made objections on various grounds to all of the principles one might expect an advocate of common-sense morality to put forward, and so it is no surprise to find him stating his overall conclusion that these principles fail to meet his four conditions (and, we may assume, fail to meet them to the point that either they can no longer plausibly be said even to be apparently self-evident or any apparent self-evidence they might retain must be admitted to be illusory) (342–3). Sidgwick draws a general connection between the clarity and non-dissensus principles, noting that attempting to make a common-sense principle precise will lead one into disagreement about it.⁵⁰ In other words, even if one becomes clear

50. For an argument that we should not expect precision in ethical principles, see Parfit forthcoming.

in one's own mind about some principle, one will find epistemic peers who disagree with one about that principle once refined. Common-sense moral principles are either too vague to meet the clarity condition, or, if they meet that condition, fail to meet the non-dissensus condition.⁵¹ We might call this the *clarity-non-dissensus dilemma (CND)*. In some cases, Sidgwick says, alternative conceptions of principles can be found in common-sense morality, and neither seems more plausible than the other (and a disjunctive principle could not be said to be sufficiently clear); in others, we cannot find a clear rule in the first place; and in yet others there are different factors in play which can be weighed against one another only by reference to the utilitarian standard (which means that the principle in question is insufficiently independent to be self-evident: see 356). Sidgwick then makes a claim about simplicity:

Even where we seem able to educe from Common Sense a more or less clear reply to the questions raised in the process of definition, the principle that results is qualified in so complicated a way that its self-evidence becomes dubious or vanishes altogether.

It might seem that Sidgwick is here introducing a further simplicity condition, in addition to his original four. In fact, however, later in the chapter he explicitly denies any tension between self-evidence and complexity, noting that a qualified principle of promising could be self-evident even if not immediately obvious, as is the definition of a circle 'as a figure bounded by a line of which every point is equidistant from the centre' (353). In other words, it is not mere complication that causes difficulties, but complication that results in unclarity and disagreement.

Sidgwick notes (343) that he is making 'a double appeal' throughout the chapter, first, to the individual reader's own 'moral consciousness', and, second, to that reader's conception of common-sense morality, as actually applied by those who she believes reliably to represent common sense. The second appeal might seem somewhat unfair to common-sense morality. The question is whether that morality can provide the *material* for self-evident principles. It is more than likely that a principle developed out of common-sense morality could not plausibly be said to be the very principle

51. Phillips (2011: 98) suggests that, when a qualified principle no longer appears self-evident, this represents a failure to meet the reflection condition. But the reflection condition seems to concern in particular the question whether some allegedly self-evident intuition is being judged on its own merits or is in fact being held as a result of custom or convention. See Schneewind 1977: 268.

actually applied within common-sense morality itself. But Sidgwick's suggestion can be taken more plausibly as putting an interpretative constraint on any reflective principle. Imagine that I propose some fairly precise principle of benevolence. I have to show not that this is the very principle actually applied by common-sense moralists, but that my principle is one of benevolence.

The rest of the chapter consists in Sidgwick's returning to the principles of common-sense morality in the order in which he has discussed them in previous chapters, and examining them in the light of his four conditions. On the whole, his examination amounts to a summary of claims already made in those chapters. He does not explicitly tie his discussion into the account of conditions for certainty, but the argument can be, and was almost certainly intended to be, interpreted in the light of them.

Often, Sidgwick confronts a common-sense principle with CND. Consider benevolence (345–6). There might be general agreement on the principle that we should benefit others to some extent. But when we try to decide who should be benefited, and by how much and in what ways, we shall find many different answers, which leads any suggested principle to fail the non-dissensus condition to the point of lacking even apparent self-evidence. But also often the problem with a principle is straightforward unclarity. Consider benevolence again. The question arises whether there is any duty actually to feel affection to others. Such affection is only partly under the control of the will, and even when it is produced through some voluntary strategy there is something unsatisfactory about it; and yet such affection is often seen, in the context of certain relationships, as a duty. In this core respect, then, the principle of benevolence cannot be stated with precision and clarity.

Further, even if the general principles of common-sense morality can be made somewhat more precise (even if, say, we could become clearer on whether there is a duty to feel affection), the practical implications of these principles are themselves unclear in particular cases. We might agree, for example, that there is a negative duty not to cause pain to others against their will and undeservedly (348). But when we consider the *limits* of such a duty—how far we may cause pain to others for their own sake, for example, or for the sake of advancing the happiness of others—the only clear principle available is the utilitarian one. This—which we might call the *under-determination objection*—is central to Sidgwick's critique of dogmatic intuitionism, and is vulnerable to the objection that Sidgwick ignores the possibility of practical judgement in particular cases.

Sidgwick's discussion of one interpretation of wisdom (343–5; see CD 6.1) brings out a further condition which any *genuinely* self-evident principle has to meet before it can be accepted—that it be substantive:

For if the rules of Wisdom and Self-control mean (1) that we ought always to do what we see to be reasonable, and (2) that we are not to yield to any impulse urging us in an opposite direction; they simply affirm that it is our duty (1) generally, and (2) under special temptations, to do what we judge to be our duty: and convey no information as to the method and principles by which duty is to be determined.

As we have seen (CD 6.1), Sidgwick then goes on to object to a substantive interpretation of wisdom, according to which it requires one to cultivate habits of rational reflection, that reason in this sense can be self-effacing, or, to use, Sidgwick's phrase, 'self-limiting'. It might be thought preferable, say, that a marriage be based on spontaneous feelings of love rather than rational design. Here it seems we have an argument appealing to consistency. The reader who is tempted by the substantive conception of wisdom is being asked to set that belief alongside other beliefs of hers about the self-limiting nature of reason which turn out to be inconsistent with it to the point that its certainty and indeed apparent self-evidence is undermined. This strategy is again quite common in 3.11. It might be tempting to think, for example, that when Sidgwick launches into detailed and substantive argument against the principle of freedom (350–1), he has broken loose of his conditions for certainty. But these arguments can themselves be seen as appeals to beliefs of the reader, and thus attempted demonstrations of inconsistency. The same interpretation can be put on arguments against a principle that it is itself inconsistent with certain other principles, as the principle of fitness is inconsistent with those of desert and of freedom (350–1).

These arguments are obviously vulnerable, since the reader may not actually possess the beliefs with which Sidgwick is urging her to compare her commitment to the moral principle under discussion. Consider, for example, the duty of parents to children (346–7). Sidgwick suggests that it is not self-evident that we owe more to our own children than to others who depend on us:

[S]uppose that I am thrown with my family upon a desert island, where I find an abandoned orphan. Is it evident that I am less bound to provide this child, as far as lies in my power, with the means of subsistence, than I am to provide for my own children? According to some, my special duty to the latter would arise from the fact that I have brought them into being: but, if so, it would

seem that on this principle I have a right to diminish their happiness, provided I do not turn it into a negative quantity; since, as without me they would not have existed at all, they can, as my children, have no claim upon me for more than an existence on the whole above zero in respect of happiness. We might even deduce a parental right (so far as this special claim is concerned) to extinguish children painlessly at any point of their existence, if only their life up to that point has been on the whole worth having; for how can persons who would have had no life at all but for me fairly complain that they are not allowed more than a certain quantity? I do not mean to assert that these doctrines are even implicitly held by Common Sense: but merely to show that here, as elsewhere, the pursuit of an irrefragable intuition may lead us unaware into a nest of paradoxes.

Sidgwick allows that someone may indeed believe in a duty to show partiality to one's own children, and then goes on to criticize the attempt to base that duty on the fact that one has brought those children into being. But there may be other grounds for the duty, including the very fact of the relationship itself. Nor, indeed, is it clear why someone who does base the duty on the fact that the parent is responsible for the child's existence may not claim that this duty is to provide the child with as good an existence as possible.

Because of the analogy he draws between legal and moral principles, Sidgwick holds a view of ethics as aiming at the level of precision appropriate to a natural science, and this explains a good deal of his antipathy towards common-sense morality.⁵² As he demonstrates so brilliantly throughout book 3, its deliverances are indeed often imprecise, unclear, or self-contradictory. Sidgwick's solution is to refer to utilitarianism to complete any common-sense principle, and his view is that such reference demonstrates that utilitarianism itself provides the foundation for common-sense morality. But, as we have seen, Sidgwick ignores or plays down the notion of practical wisdom as a capacity to judge correctly in particular cases, perhaps in the light of general 'prima facie' or pro tanto principles (see Ross 1930: 12,

52. Consider the following passage from M1 25-6: '[G]eneral rules and maxims may . . . be found mutually inconsistent . . . and here too conduct appears to us irrational, or at least imperfectly rational, not only if the maxims upon which it is professedly based conflict with and contradict one another, but also if they cannot be bound together and firmly concatenated by means of some one fundamental principle. For practical reason does not seem to be thoroughly realized until a perfect order, harmony, and unity of system is introduced into our actions.' (Quoted and discussed in Schneewind 1977: 234-5; see also Richardson 1991: 184.) At the very end of 3.11 (360-1), Sidgwick allows that common-sense morality 'may still be perfectly adequate to give practical guidance to common people in common circumstances'. But it is unsatisfactory in uncommon circumstances, and this is one reason why it fails as a system.

19–20; Hurka 2014: 135–8; CD Preface; chapter 1 n. 45). This conception of principled judgement plays an essential role in many central versions of dogmatic intuitionism, from Aristotle to Whewell (and of course on to Ross and others who wrote after Sidgwick). Sidgwick's attitude is surprising when we consider that it was Aristotle's analysis of the common-sense morality of his day that inspired Sidgwick's own approach, but less so when we reflect upon Sidgwick's brief dismissal of Aristotelian ethics at 376 for its lack of a method and his failure to take into account the role Aristotle gives to practical wisdom or *phronēsis* (see CD 4.3.2).

Sidgwick rejects 'extreme' perceptual intuitionism on the ground that most people will refer to rules at certain points (100; see CD 4.2.1). The place of aesthetic intuitionism, which allows for 'trained insight' in the context of some background rules, in Sidgwick's schema might be thought unclear, since the 'fundamental assumption' of dogmatic intuitionism is that 'we can discern certain general rules with really clear and finally valid intuition' (101). But this paragraph at 101 strongly suggests that a dogmatic intuitionist must 'remove vagueness and prevent conflict' *in particular cases* as well as at the theoretical level. So aesthetic intuitionism, for Sidgwick, is a less extreme version of perceptual intuitionism. Aesthetic intuitionism is unfavourably contrasted with the jural view of ethics, according to which agents should be told in advance exactly what their obligations are (228; see 2–4; 393; CD 1.1; 1.4). But as we have seen (CD 1.1), the law itself must allow an important place for judgement among those bound by it.

In his discussion of a pluralistic intuitionism that allows for individual judgement, John Rawls says the following:

[T]here is nothing intrinsically irrational about this intuitionist doctrine. Indeed, it may be true. We cannot take for granted that there must be a complete derivation of our judgments . . . from recognizably ethical principles. The intuitionist believes to the contrary that the complexity of the moral facts defies our efforts to give a full account of our judgments and necessitates a plurality of competing principles. He contends that attempts to go beyond these principles either reduce to triviality, as when it is said that social justice is to give every man his due, or else lead to falsehood and oversimplification, as when one settles everything by the principle of utility. The only way therefore to dispute intuitionism is to set forth the recognizably ethical criteria that account for the weights which, in our considered judgments, we think appropriate to give to the plurality of principles. A refutation of intuitionism consists in presenting the sort of constructive criteria that are said not to exist. To be sure, the notion of a recognizably ethical principle is vague, although it

is easy to give many examples drawn from tradition and common sense. But it is pointless to discuss this matter in the abstract. The intuitionist and his critic will have to settle this question once the latter has put forward his more systematic account. (1999: 34–5)

So the first question we might ask is whether Sidgwick's own utilitarianism is in a better position than pluralistic intuitionism (see Irwin 2009: 428; 491–2; Phillips 2011: 101–3; Hurka 2014: 135–7). Judgement is involved in accepting utilitarianism to start with, but it will also be involved in assessing the implications of the principle in particular cases. Consider some apparently simple case in which I have to decide between two delicious desserts. Sidgwick himself, in his discussion of empirical hedonism in 2.2–3, brings out how difficult such a decision might be. I shall have to recall earlier experiences with each kind of dessert, analyse the quality of each of those on offer, and then make a judgement based on the evidence before me. It is not clear why such judgements are any easier than trying to decide, for example, what form my expression of gratitude to some benefactor should take. Judgement is inescapable. What matters is how best to ensure that one's judgements are correct, and Sidgwick, in his discussion of the conditions for highest certainty in 338–43, provides a methodology which is at least as plausible as that of any pluralistic intuitionist.

Of course, pluralistic intuitionists should not be against system, and most are not. Consider, for example, Aristotle's discussion of how to rank the different claims on us of those differently related to us, as relatives, citizens, and so on, in *Nicomachean Ethics* 9.2, or Ross's list of prima facie duties. But we have to remember that, however much systematizing we do, the context of human decision-making is sufficiently complex and unpredictable that any plausible ethical theory has to allow some room for individual judgement about particular cases. How best to acquire that capacity for judgement is a difficult question, and one insufficiently discussed by philosophers (including Aristotle himself). But its significance is beyond reasonable doubt.

7

Egoism, Utilitarianism, and the Dualism of Practical Reason

I Egoism

1.1 A Dubious Guidance?

Egoism, for Sidgwick, constitutes an independent method of ethics in that a person may act, through weakness of will for example, against their own interests, and—according to the egoist—against a ‘dictate of reason’ (36). Sidgwick claims that egoism is widely accepted, by intuitionists and utilitarians, among others (119–20). Indeed, the majority of Christians have seen virtue as primarily the pursuit of the agent’s own happiness, and

it is hardly going too far to say that common sense assumes that ‘interested’ actions, tending to promote the agent’s happiness, are *prima facie* reasonable: and that the *onus probandi* lies with those who maintain that disinterested conduct, as such, is reasonable. (120)

Further, there is of course an intuitive foundation for egoism in the principle of prudence, which I earlier stated (CD 4.3.3) as:

P1: One ought to aim at one’s good on the whole.

Sidgwick says that this principle is ‘implied’ in rational egoism (386). It seems most likely that he sees the relation of *P1* to egoism as analogous to that of the principle of rational benevolence to utilitarianism: to complete either view, each principle requires the addition of a hedonist conception of the good.¹

1. See the final sentence of 381. Because at the very start of the *Methods* Sidgwick somewhat confusingly runs what we might call non-hedonistic or perfectionist egoism into dogmatic intuitionism (11), he equates egoism throughout the work with egoistic hedonism, rather than with the more general view that a person’s rational end is their own good or well-being (89;

Despite his accepting the ‘authority’ of self-love, Sidgwick admits that he finds egoism somewhat objectionable (199–200). Egoism sees the rules of common-sense morality as mere means to the end of individual happiness, to be ignored when self-interest requires it, and the view also fails to deliver clear practical direction: ‘A dubious guidance to an ignoble end appears to be all that the calculus of Egoistic Hedonism has to offer.’² Sidgwick notes also that there is another side to the view of egoism taken by common sense (403–4), since its focus on self-interest ‘offends both our sympathetic and our rational regard for others’ happiness’.³ The end of egoism, according to Sidgwick, does not possess the qualities we ‘divine’ (to use Aristotle’s phrase) to belong to ultimate good: it is narrow, limited, fleeting, and insecure, in comparison with the end of utilitarianism, which is vast and secure.⁴

Nevertheless, Sidgwick finds himself unable to reject the position outright:⁵

It would be contrary to Common Sense to deny that the distinction between any one individual and any other is real and fundamental, and that consequently ‘I am concerned with the quality of my existence as an individual in a sense, fundamentally important, in which I am not concerned with the quality of the existence of other individuals: and this being so, I do not see how it can be proved that this distinction is not to be taken as fundamental in determining the ultimate end of rational action for an individual. (498)⁶

95; 119). Of course, strictly egoism also requires that *P1* captures the whole truth about our reasons for action, as indeed utilitarianism requires that rational benevolence captures the whole truth; see e.g. 119; 497.

2. Hayward (1901a: 98), noting the use of ‘appears’, suggests that Sidgwick is expressing only the common view, not his own. But below Sidgwick speaks of egoism’s offending ‘our sympathetic and social nature’; see also the passage quoted next in the main text. Note that until M6, Sidgwick used the word ‘despicable’ rather than ‘ignoble’.
3. Presumably this conflict with common sense is not sufficiently ‘flagrant’ for egoism to count as ‘invalid’ (373). See Schneewind 1963: 155.
4. Sidgwick at this point (404) considers an argument of Rashdall’s (SFEC 45) that, if we require the individual to sacrifice his own happiness, we are attributing to her a different end from that which constitutes the overall good of the universe. His response is that, since the individual is special in being part of the whole, there is no problem in admitting the charge. It has not been unusual to find Sidgwick’s view on ethics described in straightforwardly utilitarian terms: see e.g. Constance-Jones 1909: esp. 927 n. And he himself sometimes identifies himself as a utilitarian: see e.g. RES 266.
5. See also e.g. P6 xx: ‘The rationality of self-regard seemed to me as undeniable as the rationality of self-sacrifice. I could not give up this conviction...’.
6. For close exegesis of this passage, see Shaver 1999: 82–98. Of the three interpretations Shaver discusses, my own is closest to the second, according to which ‘the distinction passage holds that the individual’s point of view is important and non-arbitrary’ (93) (cf. Phillips 2011: 127–8). Shaver objects that this is mere assertion. But since the distinction is fundamental, assertion is only what we should expect. Parfit finds it surprising that, given Sidgwick’s belief that the

Note that Sidgwick ascribes the view that the distinction is significant to common sense, and so it is presumably open to counter-argument like the other ‘dogmas of common sense’ (104 n) criticized by Sidgwick in book 3. But as we have seen (CD, chapter 4 n. 39), Sidgwick makes it clear in the passage which is the source of that quoted above that he finds the assumption on which ‘the rationality of Egoism is based’ to be self-evident. And Sidgwick goes on shortly after the quoted passage to note that he accepts the further common-sense view that morality’s being rational requires the demonstration of harmony between the principles of prudence and rational benevolence.

As Sidgwick elsewhere points out, a Humean may challenge egoism’s view of personal identity (418–19; see Parfit 1984: 307). On the Humean view, the ‘Ego’ is merely a system of ‘coherent phenomena’, and the egoist must therefore explain why ‘one part of the series of feelings into which the Ego is resolved [should] be concerned with another part of the same series, any more than with any other series’.

One obvious response for the egoist is, of course, to deny the Humean view itself (see CCP: 326). But even on the Humean view, it might be claimed, the ‘coherence’ between the ‘phenomena’—that is, the psychological relations between the experiences any ‘person’ has at different times—is so great that the practical implications of Humeanism are in many cases similar to those of standard egoism. And when they appear not to be, as for example in cases of advanced dementia, common sense may well follow the Humean in seeing a breakdown in egoistic rationality in such cases. In other words, the egoist may believe that Humean relations of coherence are a necessary condition of personal identity over time, or, if personal identity is not what matters, egoism may be understood as the view that the degree of concern for a future self is rational in proportion to the psychological connectedness between that future self and one’s current self (see McMahan 2002: ch. 1).

Some philosophers—utilitarians, for example—will claim that the distinction between individuals, though metaphysically robust, is ethically irrelevant. They might ask us to adopt the point of view of the universe, from which we are supposed to recognize that all our reasons rest on promoting the good overall, not some ‘part’ of it. But this is to ignore the personal point of view.⁷ As Bernard Williams puts it:

distinction between persons is fundamental, he did not give greater weight to principles of distributive justice (2011: 1.460 n. 133; see 416–17; CD 7.2.1).

7. For a defence of the irreducibility of this point of view, see Baker 2013.

[M]y life, my action, is quite irreducibly mine, and to require that it is at best a *derivative* conclusion that it should be lived from the perspective that happens to be mine is an extraordinary misunderstanding. Yet it is that idea that is implicitly contained in the model of the point of view of the universe. (1995: 170)

The model Williams has in mind must be that in which the point of view of the universe is the only point of view. But of course it may be that practical reasoning can involve more than one point of view. Consider Thomas Nagel's view:

Each person has reasons stemming from the perspective of his own life which, though they can be publicly recognized, do not in general provide reasons for others and do not correspond to reasons that the interests of others provide for him. (1986: 172)

'Stemming from' is not quite right, since it suggests that somehow the perspective, or the adoption of it, provides the reasons, perhaps through giving rise to certain self-regarding desires or concerns. But Sidgwick was right that reasons are not grounded in such motivational states. The reason *for* the special concern I have that *I* avoid severe suffering, say, is that this suffering will happen to *me*. The personal perspective is to be understood as epistemic. From it, I can grasp the normative significance of the fact that I am metaphysically distinct from others, and this will then shape my desires.

How might someone who accepts the normative significance of the personal point of view argue against a pure impartialist, who believes that the only normatively relevant perspective is that of the point of view of the universe? Mere appeal to self-evidence will of course be insufficient. Consider, then, the following example:

Torture. You are one of ten people captured in a basement by a psychopathic sadist. The sadist announces that she is going to throw a ball to each of the group in turn. Anyone who catches the ball will be put aside, to be released later. She will continue until there is only one person left, and then torture that person for many days, until death comes as a relief.

According to the impartialist, it does not matter whether you catch the ball or not. Of course, if you do not catch it, the outcome will be much worse *for you*. But it will be no worse from the point of view of the universe, since if you are not tortured someone else will be. The impartialist may be able to offer some account of *why* each of these people will think they have a reason to catch the ball, perhaps using evidence from evolutionary psychology.

Another view, however, is that they think they have this reason because they do have it.

Rational or normative egoism has a long history in philosophy. It dominated ancient ethics, and remained a serious contender throughout the seventeenth and eighteenth centuries. According to Moore (1903: 97), it began to decline in the nineteenth century because of the turn to utilitarianism; and it largely disappeared from philosophical discussion in the twentieth century perhaps in part as a result of Moore's own passionate argument, which follows the historical claim, against Sidgwick's 'absurd' view that egoism is rational.⁸ Sidgwick's position on egoism is far from absurd: he is quite right that the view deserves to be taken at least as seriously as utilitarianism and dogmatic intuitionism.⁹

1.2 Happiness and Duty

The *Methods* ends famously, and dispiritingly, with a discussion of the so-called 'dualism of the practical reason', a dualism constituted by egoism on the one hand and utilitarianism (along with dogmatic intuitionism as part of its decision-procedure) on the other. I shall discuss the dualism in the final section of this chapter. Much earlier in the *Methods*, however, Sidgwick includes within his discussion of hedonism in book 2 a fifth chapter on whether duty, as ordinarily or pre-theoretically understood, coincides with the deliverances of empirical hedonism.¹⁰ The chapter as a whole is a wonderful example of Sidgwick's insight, open-mindedness, and precision.

Sidgwick begins by distancing himself from any religious arguments on the issue (see 170), and making it clear that the view under consideration is not that any virtuous person is happier than any vicious person, but that virtue will always lead the agent to the greatest happiness possible for her in the circumstances she is in—a view often held by 'moralists of reputation' as well as 'popular preachers and moralisers' (162–3).

Most people will accept that the performance of 'self-regarding' duties tends to promote the happiness of the agent, so Sidgwick restricts the discussion to 'social' virtue, drawing Bentham's distinction between the 'external'

8. See CD chapter 3 n. 2; Crisp forthcoming.

9. In Crisp 2012b, I respond to the argument in de Lazari-Radek and Singer 2012 that Sidgwick could have avoided the dualism of practical reason had he recognized that egoism, unlike the principle of rational benevolence, can be debunked by evolutionary theory.

10. At 137–8, Sidgwick has already suggested that only in highly exceptional cases is the love of virtue inconsistent with any degree of self-concern (see CD 3.4.1).

and 'internal' sanctions (163–5). The former include legal penalties as well as the social sanctions consisting in the approval and disapproval of others. The internal sanctions consist in the pleasure of acting virtuously, the pain of remorse, and so on. It is important to note that these sanctions can conflict. In a case in which positive morality is mistaken, for example, the rules most strongly supported by the social sanctions may be different from those accepted by those with greater moral insight. Such conflicts make it less likely that duty and self-interest will coincide.

But even if we ignore them, it seems clear that the external sanctions are not sufficient to close the gap between morality and the agent's own good (165–70). In most orderly societies, open violation of the law may well be imprudent. But in circumstances of disorder, there may be great opportunities for self-advancement through vice, and even in orderly societies there will be occasions when secret crimes are egoistically rational or when immoral action is not actually forbidden by law.

Nor will the social sanctions be enough to remedy these defects. They cannot prevent secret crimes, and in rebellions or in tyrannies that result from them the criminal may be able to rely on the goodwill of those who side with her. This is especially the case with immoral acts not covered by the law. Reciprocity will indeed justify the egoist's cultivating a certain level of virtue. But what matters is appearing virtuous, so she need not refrain from secret wrongdoing. Further, some are most useful to others because of their very vices, and reciprocity anyway justifies virtue only in relation to those who can benefit one, in particular the rich and powerful. Here again, conflicting moral opinions within a society are relevant. Within some groups, special codes of duty develop which are at odds with the morality of wider society, such codes often allowing actions forbidden by morality at large. There are obvious temptations to adopt the laxer rule, and indeed doing so is often itself supported by the social sanctions (one might be thought eccentric, for example, if one follows the stricter rule).

What about internal sanctions? Sidgwick is no more sanguine here (170–5). Consider first some act of isolated duty, such as that of a soldier who undergoes a painful death. It would be 'paradoxical and extravagant' to claim that her remorse as a result of failing to sacrifice herself would make the rest of her life of negative hedonic value overall. Nor is it plausible to claim that the life of the virtuous person overall is more hedonically valuable than that of the vicious. Plato may be right that the life of those with chaotic and disordered souls is clearly less good than that of those with ordered souls. But the egoist is not

advocating disorder. Reason may well be in control of her soul; the question is what it requires.

It might be objected that an egoist cannot consult her conscience anyway, since she is already committed to the view that the only reasonable actions are those that promote her own good. But this is to forget that she may have a ‘quasi-moral’ aversion to immoral action which is not itself grounded on her view of what is reasonable and the existence of which may anyway be supported by society. Does egoism require the agent to allow these motivations to dominate in every case? This seems highly unlikely, as cases of extreme sacrifice show. It may be that in a very few cases the ‘rapture’ of virtue is so great that it does make sense to surrender oneself entirely to it, but this, as Sidgwick has already pointed out in this chapter and at 137–8, is ‘opposed to the broad results of experience, so far as the great majority of mankind are concerned’ (175).

2 Utilitarianism

2.1 Sidgwick’s Form of Utilitarianism

As now, in Sidgwick’s day the name ‘utilitarianism’ was used to refer to several quite different positions. So Sidgwick begins his fourth book with a chapter explaining how he understands the view, and how it differs from various other positions. He defines ‘Universalistic Hedonism’ (UH) as the claim that ‘the conduct which . . . is objectively right, is that which will produce the greatest amount of happiness on the whole’ (411).

By ‘objectively right’, Sidgwick means ‘right whether or not it is thought or felt to be so’ (488; see 207–8; 344 n. 2; 394; 429–30).¹¹ An action’s being subjectively right, then, consists in the agent’s believing it to be right. Now consider the following case.

The Rash Doctor. You are suffering from some painful medical condition, for which two drugs are available. Drug A might cure you completely, and there is a 1 per cent probability of its doing so. But there is a 99 per cent probability of its killing you. There is a 100 per cent probability that drug B will almost cure you (though it will leave you with a very slight twinge of pain, once every year or so). Your doctor, in full awareness of these facts, prescribes drug A, and it cures you completely.

11. For further discussion of different senses of ‘right’ and ‘wrong’, see Parfit 2011: ch. 7.

On Sidgwick's view, it seems, not only is the rash doctor's action objectively right, but the prescription of drug B would have been objectively wrong. This may seem counter-intuitive, but we should remember that Sidgwick keeps the notions of wrongness and blameworthiness distinct. Blaming, as an activity, itself falls within the scope of UH (221; 491; 499), and it is clear that blaming doctors who take such risks with their patients will be required. Sidgwick sometimes uses the term 'reasonable' as an alternative to 'right', however, and the rash doctor's action seems to be a paradigmatic case of acting unreasonably. But this use of 'reasonable' is that found within common-sense morality, which is again governed by UH. What Sidgwick has in mind as the reasonable action is that which the agent has most reason to perform, and that action is the one which will in fact produce the greatest good. In other words, the ideal agent, according to UH, will be the one who always performs actions that maximize happiness overall. The question of which strategy will bring each of us closest to that ideal is, of course, a difficult matter, and much of book 4 is concerned with it.

In contemporary philosophy, 'subjective rightness' for a utilitarian is usually understood in terms of *expected* utility, that is, the value of some outcome multiplied by its probability of coming about. And utilitarian theories themselves are usually stated in terms of such rightness, so that the right action for the rash doctor would have been to prescribe drug B. Sidgwick shows some awareness of the importance of probability. We judge the value of some possible pleasure partly in the light of its certainty or uncertainty, and any 'doubtfulness . . . detracts from its value' (124 n. 1; see 123–4 n; 383 n). Empirical hedonism requires us to judge lines of conduct 'taking all probabilities into account' (131; see 142), and

we have in each case to compare all the pleasures and pains that can be foreseen as probable results of the different alternatives of conduct presented to us, and to adopt the alternative which seems likely to lead to the greatest happiness on the whole.¹² (460)

If 'likely' here means 'most likely', then Sidgwick seems to be claiming that we are required to maximize the probability of our bringing about the greatest happiness.¹³ This view runs into the same sort of problem as that based on objective rightness, since it may require us to prefer an action with only the

12. Note also Sidgwick's suggestion that intention concerns all consequences 'foreseen as certain or probable' (202).

13. A structurally similar claim is made in the first sentence of the main text of Kagan 1989.

tinest chance of producing the best possible outcome over an action which is certain to produce an outcome only slightly less good than the best possible. I have little doubt that Sidgwick would accept that he intended here to speak of something like expected utility.¹⁴

Nevertheless, Sidgwick has nothing like the modern conception of subjective rightness. Probabilities enter only at the level of decision-procedures, and the choice of any decision-procedure will be objectively right only to the extent that it produces the best outcome overall. And once we recognize both the distinction between objective wrongness and blameworthiness, and that between the criterion of rightness and decision-procedures, Sidgwick's view seems at least as plausible as modern accounts based on expected utility.

Before further clarifying UH, Sidgwick distinguishes it from three other views (411–12). The first is egoistic hedonism, and here Sidgwick refers back to his explanation of the close connections often alleged between the two forms of hedonism (84–5):

Among other points it was there noticed that the confusion between these two ethical theories was partly assisted by the confusion with both of the psychological theory that in voluntary actions every agent does, universally or normally, seek his own individual happiness or pleasure. Now there seems to be no *necessary* connexion between this latter proposition and any ethical theory: but in so far as there is a natural tendency to pass from psychological to ethical Hedonism, the transition must be—at least primarily—to the Egoistic phase of the latter. For clearly, from the fact that every one actually does seek his own happiness we cannot conclude, as an immediate and obvious inference, that he ought to seek the happiness of other people. (412)

Sidgwick is quite right that the highly significant distinction between egoistic and universalistic forms of hedonism had not been drawn sufficiently clearly in earlier literature. Nevertheless, in his earlier discussion, there is no mention of any confusion between egoistic or universalistic hedonism and (egoistic) psychological hedonism. Rather, Sidgwick notes that Bentham and those of his 'disciples' who accepted psychological hedonism sought to 'base' UH on psychological hedonism. Sidgwick mentions in particular Mill's proof. His objection to that, however, is not that it confuses egoistic or universalistic hedonism with psychological hedonism, but that, because there is no actual desire for the general happiness, the general happiness cannot be

14. And probably Kagan too (see previous note).

proved to be desirable by Mill's own argument (388). There is also a possible tension here in Sidgwick's argument. The initial suggestion is that because both egoistic and universalistic forms of hedonism are confused with egoistic psychological hedonism, egoistic and universalistic forms of hedonism are themselves confused with one another. But, Sidgwick goes on to suggest, if one starts as an egoistic psychological hedonist, the natural ethical or normative direction in which one might move would be towards egoistic rather than universalistic hedonism. This appears to imply that it is more likely that the two egoistic forms of hedonism will be confused with one another than that universalistic hedonism will be confused with (egoistic) psychological hedonism. And it is anyway highly implausible to think that people confuse universalistic hedonism with (egoistic) psychological hedonism.

Sidgwick goes on to offer two further clarifications. Utilitarianism is independent of the psychological view that the moral sentiments derive from non-moral pleasures and pains caused by different kinds of conduct. And we must remember that UH is a theory about rightness of action, and not goodness of motive. UH may well direct us to prefer motives other than pure impartial altruism on the ground that they are more productive of happiness overall.

The second section of 4.1 discusses the scope of utilitarianism, and its relation to equality, foreshadowing debates concerning the nature of egalitarianism and the moral status of non-human animals and future generations that became central in the last few decades of the twentieth century. Sidgwick begins by repeating his assumption that pleasures and pains must be commensurable, allowing for a certain degree of vagueness: 'each may be at least roughly weighed in ideal scales against any other' (413; see 416). He then claims, plausibly enough, that it would be 'arbitrary and unreasonable' not to include the happiness of non-humans within the scope of utilitarianism, noting that the difficulty of assessing non-human experience will arise for any ethical view which does not ignore it. At this point, Sidgwick goes on to make a highly significant breakthrough in our understanding of utilitarianism. Having noted that UH must be temporally neutral, since the value of happiness does not depend on when it exists, Sidgwick asks what the implications of utilitarianism are regarding population size. It might be thought (and, Sidgwick says, Malthusians indeed have thought) that we should aim to maximize average happiness. But this is to ignore the fact that adding extra members to a population, if their lives are of positive value,

may increase the total level of happiness even if the average happiness in the new population is lower than that in the original.¹⁵

The only potential absurdity in this position noted by Sidgwick is its apparent commitment to exactness of measurement of the value of happiness. He fails to mention that a readiness to sacrifice quality or intensity of well-being for a greater overall total brought about by increased temporal quantity results in implications such as Parfit's *Repugnant Conclusion*:

For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better even though its members have lives that are barely worth living. (Parfit 1984: 388)¹⁶

One way to avoid the Repugnant Conclusion is through the notion of discontinuity of value, according to which, at some point, increases in the temporal quantities of well-being can no longer outweigh losses of quality. But Sidgwick denies discontinuity (123–4 n; see CD 3.4) and it is tempting to think that he would have seen the Repugnant Conclusion as just another implication of utilitarianism counter to the morality of common sense but not to properly grounded ethical theory.

The chapter ends with an interesting discussion of how utilitarians might choose between different distributions of the same amount of happiness. As Sidgwick notes, this question is not purely academic, since the vagueness of actual hedonistic calculations will often result in our being unable to see any quantitative difference between two or more distributions, even if in fact there is one. As Sidgwick points out, utilitarianism will be indifferent between any set of distributions, all of which maximize happiness overall. And so:

we have to supplement the principle of seeking the greatest happiness on the whole by some principle of Just or Right distribution of this happiness. The principle which most Utilitarians have either tacitly or expressly adopted is that of pure equality—as given in Bentham's formula, 'everybody to count for one, and nobody for more than one'. And this principle seems the only one which does not need a special justification; for, as we saw, it must be reasonable to treat any one man in the same way as any other, if there be no reason apparent for treating him differently. (416–17)

15. This insight is brilliantly expounded and developed in part 4 of Parfit 1984.

16. The 'original' repugnant conclusion can be found in McTaggart 1927: 453–4. (McTaggart himself does not find the conclusion repugnant.)

This brief but fascinating passage raises several questions. First, why does Sidgwick feel entitled to describe Bentham as an egalitarian? Second, why does he see a need for a supplementary principle at all? It is not as if the utilitarian answer in such cases is unclear: any distribution that maximizes is acceptable (see Broad 1930: 251; Irwin 2009: 514–16). Is Sidgwick again feeling the need for comprehensive ethical guidance? Third, if this principle of equality is itself independent of the utilitarian principle, why does it come into play only as a tiebreaker?¹⁷ Why can it not, like egoistic hedonism, be found in practical conflict with the utilitarian principle? Consider the following pairs of possible distributions:

Pair I

(a)

<i>Person A</i>	<i>Person B</i>
-----------------	-----------------

50	50
----	----

(b)

<i>Person A</i>	<i>Person B</i>
-----------------	-----------------

1	99
---	----

Pair II

(a)

<i>Person A</i>	<i>Person B</i>
-----------------	-----------------

50	50
----	----

(b)

<i>Person A</i>	<i>Person B</i>
-----------------	-----------------

1	99.01
---	-------

On Sidgwick's view, the Benthamite equality principle requires us to choose (a) over (b) in Pair I. But it has no force in Pair II, where we have to choose (b) over (a) on utilitarian grounds. But if equality has no weight against even the smallest amount of utility, one cannot help but wonder why it should have any weight in choices between distributions of equal utilitarian value.

Another question is to which earlier passage or passages Sidgwick is referring when he says 'as we saw'. Later in book 4, when Sidgwick is discussing impartiality, he describes it as 'a special application of the wider

17. See Ross 1939: 72. At 441–2 and 447, Sidgwick repeats the claim that utilitarianism is committed to this kind of equality. In the latter passage, he notes that such a principle is likely to maximize happiness in practice, since people dislike 'unreason' and have an aversion to inferiority to others.

maxim that it cannot be right to treat two persons differently if their cases are similar in all material circumstances', and refers back to 3.13.3, going on to claim that utilitarianism 'as we saw, admits this maxim no less than other systems of Ethics' (441–2). This 'as we saw' must refer back to our passage, so it is in 3.13.3 that we must look. Given that the egalitarian principle requires no justification, it is tempting to think that it is evident in itself, and the most plausible candidate is what Sidgwick calls the principle of 'Justice', which is stated as follows:

'if a kind of conduct that is right (or wrong) for me is not right (or wrong) for some one else, it must be on the ground of some difference between the two cases, other than the fact that I and he are different persons.' A corresponding proposition may be stated with equal truth in respect of what ought to be done *to*—not *by*—different individuals. (379)

But this principle will not support a substantive principle of equality of the kind Sidgwick is advocating.¹⁸ Consider the following pair of distributions:

Pair III

(a)		
<i>Person A</i>		<i>Person B</i>
99		1
(b)		
<i>Person A</i>		<i>Person B</i>
1		99

If someone were to choose (a) over (b) merely on the ground that Person A is different from Person B, she would be open to objection in the light of the Justice principle. But this objection does not apply to the utilitarian who chooses (i.e. 'plumps for') (b) over (a) in Pair I above. She sees the difference between persons as ethically irrelevant, and so sees no relevance either in the fact that Person A is different from Person B or in the fact that Person A does much worse in (b) than she does in (a).

18. See Hayward 1901a: 140; Locke 1968. As Hayward notes (143), Sidgwick himself admits (380) that the effect of the principle, 'strictly speaking, is merely to throw a definite *onus probandi* on the man who applies to another a treatment of which he would complain if applied to himself'. It might be thought that this is enough to justify equality. Imagine a utilitarian distributor who chooses (perhaps randomly) (b) over (a) in Pair I. Couldn't Person A object that the distributor would complain were she to end up with one unit rather than fifty? Hardly. The distributor is a utilitarian, so would not recognize the complaint. (She might not *like* being at level 1, but that is a different matter.)

2.2 The Arguments for Utilitarianism

In seeking Sidgwick's argument for utilitarianism, we find ourselves again in 3.13.3. The prospects for an intuitionist grounding for utilitarianism appear somewhat bleak at the beginning of this section, since Sidgwick tells us that the principles he is to outline 'are of too abstract a nature, and too universal in their scope, to enable us to ascertain by immediate application of them what we ought to do in any particular case; particular duties have still to be determined by some other method' (379).¹⁹ Later, however, he writes that the 'axiom of Rational Benevolence is . . . required as a rational basis for the Utilitarian system' (387). That axiom is:

*B*₃: 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, or less certainly knowable or attainable by him. (See CD 4.3.3.)

If this principle is understood monistically, and not as one among several others in some version of dogmatic intuitionist pluralism, it might appear to provide not just a basis for utilitarianism, but a statement of it.²⁰ Sidgwick, however, as we have seen, prefers to define utilitarianism as a version of hedonism, so for *B*₃ to constitute a statement of utilitarianism would require substituting 'happiness' for 'good'. And it is that substitution which is required for *B*₃ to become sufficiently concrete for it to constitute a

19. Recall Sidgwick's definition of a 'method' in the first sentence of the book: 'any rational procedure by which we determine what individual human beings "ought" . . . to do'. On their own, the abstract principles are insufficiently determinate to count as a method. See CD 1.4.
20. See Deigh 2010: 82. Schneewind (1977: 306–7) finds problematic the move from *B*₃ (which is also his *B*₃) to 'the fuller principle *B*₅' ('it is right and reasonable for me to do what I believe to be ultimately conducive to universal Good or Happiness' (Schneewind 1977: 297; adapted from M7 507)). He suggests that (1) the notion that goodness provides a reason must be added; (2) the notion of maximization must be incorporated; and (3) the idea that an act's maximizing goodness is not just an indication of its being right but the right-making characteristic must be included. I take it that by 'regard' Sidgwick means more than merely 'contemplate'. Consider his use of the notions of 'self-regard' (e.g. P6 xx (contrasted with 'self-sacrifice'); 48; 204), 'regard to' (e.g. 204 (a regard to happiness is identified with 'self-love')), and 'regard for' (e.g. 204 (seen as a motive); 326 (regard for the well-being of others is kindness); 372 (regard for social well-being is a motive); 502 (regard for the general good is utilitarian duty, which demands sacrifice)). The fact that *B*₃ is an elucidation of benevolence, which Sidgwick clearly sees as a practical virtue, also suggests that Sidgwick has a normative sense of 'regard' in mind here. It is true that maximization must be taken to be implicit, and is made explicit only when Sidgwick defines utilitarianism itself (see GSM 110; Irwin 2009: 498–9). (Note that elsewhere Sidgwick suggests that the right action will be that which brings about the *greatest* good (113), and that egoism advocates the agent's *greatest* happiness as ultimate end (119).) But we can presumably take *B*₃ to be identifying a right-making characteristic, since it states an ultimate normative principle.

practical method of ethics (388; 421 n. 1; see de Lazari-Radek and Singer 2014: 121). How will that substitution be carried out? The reference to ‘some other method’ in the passage quoted above from 379 might suggest that it will not be through philosophical intuitionism, as also the closing sentence of 3.13: ‘The identification of Ultimate Good with Happiness is properly to be reached, I think, by a more indirect mode of reasoning; which I will endeavour to explain in the next Chapter’ (389).²¹

Consider, however, Sidgwick’s clinching move in his argument for hedonism in 3.14. Having set out the view that the ‘conscious life’ which constitutes the ultimate good includes not only pleasure or happiness but also the ‘ideal goods’ of cognition of truth, beauty, and free or virtuous action (400), Sidgwick continues:

I think, however, that this view ought not to commend itself to the sober judgment of reflective persons. In order to show this, I must ask the reader to use the same twofold procedure that I before requested him to employ in considering the absolute and independent validity of common moral precepts. I appeal firstly to his intuitive judgment after due consideration of the question when fairly placed before it: and secondly to a comprehensive comparison of the ordinary judgments of mankind. (400; see 343)

Sidgwick’s hedonism, then, is justified in the same way as the principles of justice, prudence, and benevolence, through a foundationalist appeal to intuition (see McTaggart 1906: 408; Frankena 1956: 539–44; Skelton 2008: 191–3; CD 3.3.3; cf. Phillips 2011: 97; 102; III n. 7; de Lazari-Radek and Singer 2014: 146 n. 46), supported by an Aristotelian appeal to the common sense of humanity. So what is meant by his references in 3.13 to ‘some other method’ and ‘a more indirect mode of reasoning’? Much of the earlier sections of 3.14 are taken up with ground-clearing, with various views being disposed of: that the good consists in virtue (3.14.1), in acting subjectively rightly (3.14.2), in what promotes survival (3.14.3), and so on. Sidgwick perhaps felt that hedonism, unlike the other three principles, would not appear *immediately* self-evident,²² so some non-intuitionist,

21. The contrast is with the ‘directness’ of Mill’s argument to the conclusion that happiness alone is desirable from the premise that it alone is desired. But we still need to understand why a (direct) appeal to intuition on its own will not be enough.
22. At CIM 25, Sidgwick claims that most ‘plain men’ would say that they saw the obligations of Justice and Benevolence ‘clearly and immediately’ (in this piece, Sidgwick ignores egoism). They would also say that they saw the principles of veracity and so on equally clearly and immediately; hence, the need for the detailed examination of 3.3–10. In other words, the argument for the three principles of 3.13 has also been indirect, since it has required the ground-clearing of the earlier chapters in book 3.

straightforward philosophical argument would be required to provide the right conditions for intuitionist reflection and Aristotelian dialectic to do their work.²³

Sidgwick begins 4.2 ('The Proof of Utilitarianism') by noting that utilitarianism appears to stand in special need of proof, or what Mill calls 'considerations determining the mind to accept it'.²⁴ Common sense takes it as read that individuals have reasons for promoting their own good, and that certain moral rules are binding, so is disinclined to question egoism or dogmatic intuitionism. Further, since utilitarianism claims to be a 'higher' principle than both egoism and dogmatic intuitionism, the proponents of each will naturally require some account of its legitimacy.²⁵

Sidgwick accepts that any proof that attempts to infer utilitarianism, as a 'first principle', from certain premises will fail, since these premises themselves will be the first principles. By a principle's being first, then, Sidgwick means its being non-inferential—an important part of what it is to be self-evident. If a utilitarian is to prove her principle (inferentially) to someone who accepts certain other principles, 'the process must be one which establishes a conclusion actually *superior* in validity to the premises from which it starts'. I take it that by 'validity', here, Sidgwick means something like 'credibility'. So, when there is a conflict between the principles of common-sense morality or the requirements of rational egoism, which the intuitionist and the egoist respectively took to be self-evident, and utilitarianism, the (now ex-)intuitionist and (ex-)egoist will find utilitarianism more persuasive: 'Utilitarianism... must be accepted as overruling Intuitionism and Egoism.'²⁶

23. Another possibility is that Sidgwick has in mind the arguments at 401–6, which are not so much arguments for hedonism as arguments to explain away the view that there are non-hedonic goods. Derek Parfit has suggested to me that Sidgwick may have had in mind his claim that hedonism provides a means of commensuration (406–7; see CD 3.3.3).

24. This appears to be a paraphrase of Mill 1998: 1.5.20–2.

25. It might seem that there is a tension here with Sidgwick's view that the utilitarian method is itself implicit in common sense (14). Presumably what he means is that a utilitarian has to justify not the principle of benevolence, but the replacement of all other ultimate moral principles with that one alone.

26. Schneewind (1977: 263) interprets validity here non-epistemically, such that for one principle to be superior in validity to another is just for it to have greater normative authority (rather than greater credibility, certainty, or whatever). But Sidgwick ascribes 'superior validity' initially not to a moral principle itself, but to a conclusion of an argument in relation to its premises (even if that conclusion is a moral principle). In the *Methods*, Sidgwick nearly always uses 'validity' in an epistemic context, though at least once he uses the term to refer to the applicability of common-sense moral rules within a utilitarian system (8). See CD, chapter 1 n. 10.

In other words, intuitionist or egoist principles have to be given some epistemic weight in the argument, or the proof will not be addressed to intuitionists or egoists. So Sidgwick proposes ‘a line of argument which on the one hand allows the validity, to a certain extent, of the maxims already accepted, and on the other hand shows them to be not absolutely valid, but needing to be controlled and completed by some more comprehensive principle’ (420).²⁷ At this point, Sidgwick does refer back to 3.13, claiming that he there presented a ‘controlling and completing’ (C&C) argument addressed to the Egoist (420).²⁸ The beginning of the following paragraph shows that the passage to which Sidgwick is referring is that in which he argues for B_3 . He probably has in mind the statements of B_1 and B_2 (382), though nothing in this passage or its context suggests that the argument is being addressed to the egoist in particular.²⁹ It can be read as such, however, as Sidgwick goes on to show:

When . . . the Egoist puts forward, implicitly or explicitly, the proposition that his happiness or pleasure is Good, *not only* for him but from the point of view of the Universe—as (*e.g.*) by saying that ‘nature designed him to seek his own happiness’—it then becomes relevant to point out to him that *his* happiness cannot be a more important part of Good, taken universally, than the equal happiness of any other person. And thus, starting with his own principle, he may be brought to accept Universal happiness or pleasure as that which is absolutely and without qualification Good or Desirable: as an end, therefore, to which the action of a reasonable agent as such ought to be directed. (420–1; see 497–8)

This is a C&C argument in that the now-converted egoist may retain the thought that she has reason to concern herself with promoting her own good, but only because so doing will advance the universal good.

If Sidgwick were a pure foundationalist, we could understand C&C arguments as strictly tactical. Their aim would be to bring one’s opponent, through reflection on the principles she herself accepts, to see that lying

27. By ‘absolutely valid’ here, I take Sidgwick to mean something like ‘credible in themselves as stating the full truth about our reasons for action, without reliance on inference’. See 400, where Sidgwick speaks of ‘absolute and independent validity’; I take the ‘and’ here to be, in effect, expegetic.

28. See EEFP 30; Schneewind 1977: 288.

29. Indeed, there is evidence in the first edition that this is how Sidgwick saw the matter: ‘this line of argument is addressed to the Intuitionist also: in fact it was presented in ch. 13 of the preceding book as one of the few products of the Intuitionist method which stood the test of rigorous criticism, and it was shewn that it occupied a prominent place in the exposition of Ethics by the most unqualified and confident advocates of this method’ (M1 392).

beneath those principles is some more fundamental and self-evident principle, which itself explains any plausibility that the original principles might have.³⁰ There is no doubt that this rhetorical role is indeed one that Sidgwick ascribes to C&C arguments. But in fact, they have weight in themselves, as part of the process of Aristotelian dialectic which Sidgwick runs—somewhat unstably, I have suggested (see CD 4.2.3)—alongside the methodology of philosophical intuitionism.³¹ In other words, Sidgwick is granting ‘validity, to a certain extent’ to common-sense morality because he does indeed believe it to have credibility independent of any inference from a self-evident principle.³² As he later says: ‘[T]he Utilitarian argument cannot be fairly judged unless we take fully into account the cumulative force which it derives from the complex character of the coincidence between Utilitarianism and Common Sense’ (425; see 373). As far as dogmatic intuitionism is concerned, the utilitarian has first to demonstrate that the principles of veracity, justice, and so on have merely ‘a dependent and subordinate validity’, in that *some* higher principle is required to explain exceptions as well as resolve indeterminacy and conflict (421–2). This ‘negative’ aspect of

30. Note the reference to ‘Aristotle’s distinction between logical or natural priority in cognition and priority in the knowledge of any particular mind’ at EEFP 29 (see Schneewind 1977: 287). This position is close to the second solution to the puzzle raised by Phillips (2011: 65–76) about why Sidgwick offers C&C arguments (a form of what Phillips calls ‘bipartite arguments’) in book 4 without referring back to the arguments based on self-evidence in 3.13 (which Phillips calls ‘criterial’ arguments). Phillips’s own solution is to claim that Sidgwick offers the book 4 arguments because he sees that the book 3 arguments have failed, as Sidgwick admits in connection with Egoism at 420–1. This seems to me unlikely. At 498, Sidgwick suggests that, if a harmony between the maxims of Prudence and Benevolence can be established, morality will be ‘completely rational’. That harmony, of course, he cannot find, but this is where the failure occurs, not in the failure of the Utilitarian to persuade the Egoist. Understood in the right way, they could both be correct, as I shall suggest (CD 7.3). Phillips also suggests (2011: 70) that in book 4 Sidgwick describes the earlier, criterial arguments as merely *part* of the later bipartite arguments, and not as having any independent weight. But the passage cited by Phillips (M1 392) concerns not the arguments of 3.13 but the earlier arguments of book 3, to the conclusion that the principles of common-sense morality ‘have only a dependent and subordinate validity’. Nevertheless, as I shall suggest in the main text, the criterial arguments are in the background in 4.2, since one of Sidgwick’s hopes is that the bipartite arguments of 4.2 will enable his opponent to see the force of the criterial arguments. But this is no evidence that Sidgwick sees criterial arguments as having no independent force.
31. See Phillips’s third solution (2011: 73). Phillips explains away the key passage on 373 as exemplifying a confusion between the ‘epistemic [philosophical] and moral-theoretic [dogmatic] components of intuitionism’. But the fact that the passage, placed as it is at the opening of one of the most significant chapters of the book, survived several of Sidgwick’s careful revisions makes Phillips’s suggestion especially implausible.
32. At the conclusion of M1 (472), Sidgwick says that he has established benevolence as the foundation of duty ‘without impairing our confidence in the substantial veracity of Common Sense’.

the argument Sidgwick suggests he has ‘sufficiently developed’ in book 3, and continues:

It remains to supplement this line of reasoning by developing the positive relation that exists between Utilitarianism and the Morality of Common Sense: by showing how Utilitarianism sustains the general validity of the current moral judgments, and thus supplements the defects which reflection finds in the intuitive recognition of their stringency; and at the same time affords a principle of synthesis, and a method for binding the unconnected and occasionally conflicting principles of common moral reasoning into a complete and harmonious system. If systematic reflection upon the morality of Common Sense thus exhibits the Utilitarian principle as that to which Common Sense naturally appeals for that further development of its system which this same reflection shows to be necessary, the proof of Utilitarianism seems as complete as it can be made. And since, further—apart from the question of proof—it is important in considering the method of Utilitarianism to determine exactly its relation to the commonly received rules of morality, it will be proper to examine this relation at some length in the following chapter. (422)

The difference between the negative and the positive relation of utilitarianism to the morality of common sense is not straightforward.³³ When he is spelling out his conception of the negative phase of the argument, Sidgwick makes no reference to utilitarianism itself, though the criticism of common-sense morality he imagines being that of a utilitarian, and he says immediately beforehand that *utilitarianism* has to show itself in both the negative and positive relation to the principles of common-sense morality. Further, at 423, he again mentions the twofold relation, and refers back to 86, where the negative phase of argument includes not only criticism of common-sense morality but common-sense morality’s supersession by utilitarianism. Nevertheless, Sidgwick’s argument is most coherent if the negative phase is understood to be purely critical. The positive phase then consists in ‘showing how *Utilitarianism* sustains the general validity of the current moral judgments’, and so on (422; my italics).

33. Schneewind’s renaming of the distinction of the negative and positive arguments, presumably in the light of 421–2, as the ‘dependence’ and the ‘systematization’ arguments (1977: 264–5) may lead to confusion, since the positive (systematization) argument is, among other things, aimed at showing that it is utilitarianism on which common-sense morality depends.

As I have noted, Sidgwick claims that he has ‘perhaps sufficiently developed’ the negative phase of the argument in book 3. In the second edition, the phrase ‘*aliud agens*’—‘while pursuing another goal’—was removed. This may be because he recognized that much of the critical analysis of common-sense morality was in fact already being carried out from a utilitarian perspective. As we have seen (CD 1.1), Sidgwick avows a position of neutrality throughout his investigation of common-sense morality. But even if he succeeds in procedural neutrality, it clearly cannot be said that the outcome of his enquiry is neutral: as far as dogmatic intuitionism’s claim to being the correct account of the fundamentals of ethics is concerned, it is negative. Nevertheless, the negative/positive distinction plays no official role in setting the context or constructing the argument of book 3, and this gives rise to another problem with the ordering of Sidgwick’s argument. If book 3 were purely negative, no mention would have been made of utilitarianism as the higher principle needed to supersede the principles of common-sense morality.³⁴ But utilitarianism is in fact mentioned fairly often in just such a role, which has the consequence that book 3 involves the positive as well as the negative phase of the utilitarian argument against common-sense morality.³⁵ This means that the lengthy chapter, 4.3, in which Sidgwick presents the positive phase of the argument covers ground which has already been covered once, or indeed more than once, in the discussions of individual virtues in 3.3–10 and then the review of common-sense morality in 3.11. His reason for doing so is presumably, by focusing in more closely on the detail of the consistency and coherence of the relation between utilitarianism and common sense, to bring out the ‘cumulative force’ mentioned in the passage cited above from 425. But if I am right about Sidgwick’s reason for deleting ‘*aliud agens*’, he himself might have come to the view that it would have been wiser to consider the relation between utilitarianism and common-sense morality fewer than three times.

34. Though of course we should expect frequent reference to the principle of benevolence, as one principle among others, and it is important to remember that common-sense morality has this resource available to resolve certain conflicts and unclaritys. See Shaver 2011b: 584–6.
35. Sidgwick himself shows awareness of the blurring of the negative/positive distinction. At 355, for example, he claims that it was made clear in 3.8 (that part of the negative argument concerned with minor social duties) that common sense can avoid inconsistency or vagueness only ‘by adopting the “interest of society” as the ultimate standard’ (which is of course the main claim of the positive argument). See also e.g. 302–3.

2.3 Utilitarianism and Common-sense Morality

Sidgwick begins 4.3 by citing earlier writers—Cumberland, Shaftesbury, Hume, and Smith—who have drawn attention to the utility of moral virtues, concluding that common-sense morality ‘may be truly represented as at least unconsciously Utilitarian’ (424; see Gibbard 1982).³⁶ But, as he goes on to point out, saying that common-sense morality is generally useful or beneficial is one thing, and asserting utilitarianism as the true theory of the criterion of rightness another (see Simmons 1982: 95; 97). Again, Sidgwick tells us that a C&C argument is required:

Utilitarians are rather called upon to show a natural transition from the Morality of Common Sense to Utilitarianism, somewhat like the transition in special branches of practice from trained instinct and empirical rules to the technical method that embodies and applies the conclusions of science: so that Utilitarianism may be presented as the scientifically complete and systematically reflective form of that regulation of conduct, which through the whole course of human history has always tended substantially in the same direction. (425)

The thought seems to be, then, that if the intuitionist can be brought to see that existing moral rules have developed because of their general usefulness, she will then—presumably also recognizing the deficiencies of those rules brought out in the negative phase of the argument—grasp that universal benevolence is the only one of those principles she formerly took to be self-evident which survives reflection as both self-evident in itself and also supported by dialectical considerations, including its consistency with common-sense morality itself.³⁷ The main problem for Sidgwick is that the dogmatic intuitionist can accept nearly everything Sidgwick says about the general usefulness of common-sense rules and dispositions, but deny the move to utilitarianism as the correct account of the criterion of rightness (consider, for example, Ross’s view of promissory obligation as binding in itself, independently of the usefulness of promise-keeping in any particular

36. Note that Sidgwick appears to be assuming that the only consequences that matter to common sense (unconsciously) are hedonic: see Irwin 2009: 474.

37. Demonstrating the degree of this consistency is Sidgwick’s primary task in 4.3, since it is this that provides ‘cumulative force’. Utilitarianism (i) supports the rules of common sense; (ii) justifies their qualifications; (iii) explains apparent anomalies; (iv) solves problems of indeterminacy; (v) supports the standard view of the relative importance of different duties; (vi) resolves conflicts between rules, and between individuals who interpret those rules differently; and (vii) explains the disparities between moral codes at different times and in different places (425–6).

case; see Irwin 2009: 447; 478).³⁸ Sidgwick does not, then, have a watertight argument for utilitarianism. But moral theories are here on a par, and Sidgwick is surely right to think that at least some of his readers may be prompted to make the move to utilitarianism after reflection on the usefulness of common-sense morality.

A similar problem arises with a further, psychological argument for the harmony of common sense with utilitarianism, discussed by Sidgwick at 461–2. According to the hypothesis underlying this argument, moral sentiments are derived from experiences of pleasure and pain: a human being dislikes actions that cause pain to herself or anyone to whom she has some special relation; she will be restrained from performing actions that cause pain to others through fear of their resentment, and through sympathy with others. An analogous story can be told about actions that cause pleasure. The similar likes and dislikes that develop in a community are then transmitted by tradition and imitation, and become what are now known as the moral sentiments.

Sidgwick himself rejects this as a full account of the origin of moral judgement, since the deliverances of his own moral consciousness are principles too abstract to have any noticeable relation to particular experiences, and continues:

But that the theory gives a partially true explanation of the historical origin of particular moral sentiments and habits and commonly accepted rules, I see no reason to doubt; and thus regarded it seems to supplement the arguments of the preceding chapter that tend to exhibit the morality of common sense as unconsciously or ‘instinctively’ utilitarian. (462–3)

Again, though there is nothing to prevent a dogmatic intuitionist from accepting the psychological account and continuing to deny utilitarianism, some may be led to accept the truth of utilitarianism through reflection upon the origins and functions of common-sense, or positive, morality.

There is, however, one apparently major difference between common-sense morality and utilitarianism at the most general level, a difference of which earlier writers—including Hume, Smith³⁹, and Sidgwick himself—were

38. ‘If, so far as I can see, I could bring equal amounts of good into being by fulfilling my promise and by helping some one to whom I had made no promise, I should not hesitate to regard the former as my duty . . . [and] normally promise-keeping . . . should come before benevolence’ (Ross 1930: 18–19). Ross, then, appears committed to the principle that any agent has an ultimate reason to keep promises.

39. Hume 2007 3.3.5.6; 1998: 5.1 n. 17; Smith 1976: 4.2.4.

well aware, but which has not received the attention it should in more recent philosophical ethics. This is the very notion of moral value itself. Why, Sidgwick asks in 426–7, do we feel that the excellence or value of virtue is so different from that of, say, a machine, or of the beauty or intellect of some human being?

Sidgwick's answer (and here he refers back to 3.2 [220]) is that, since virtue, strictly understood, is always thought to be to some degree voluntary, the main obstacle to virtue is lack of motive. So we expect that moral judgements will have a positive motivational effect, and our awareness of this fact explains the degree of difference between the experience of a moral sentiment and the pleasure or pain we feel when contemplating non-human, or human but non-voluntary, utilities or disutilities.

Sidgwick's argument appears to be that, because we know that the moral judgements we make might actually make a difference to the world through motivating an agent, this provides a special positive or negative frisson not found in other evaluative judgements. This is an interesting but somewhat problematic suggestion. First, Sidgwick seems to be putting excessive weight on cases in which we are voicing moral judgements in an attempt to motivate someone either to act or to refrain from acting in some way. But often we voice such judgements in non-motivational contexts, such as assessments of past behaviour, or indeed keep our judgements to ourselves. Second, the phenomenology of such judgements suggests that, on the face of it, they are responses to admirable or objectionable qualities, in a special category, and that we have the responses we do—morally admiring, blaming, and so on—because of the nature of those qualities themselves.

Sidgwick goes on to raise an objection to his own argument: that there are dispositions to voluntary action which are more useful than many virtues but not viewed as such. He cites several examples from Lecky: self-interest, a disposition to take revenge, and unscrupulous statesmanship. Having reminded us that he is not claiming a complete overlap between utilitarianism and common-sense intuitions, Sidgwick offers several responses. First, we must distinguish the goodness of dispositions from the rightness of conduct. So a utilitarian may condemn a particular action which results from a general disposition—say, conscientiousness—which has utilitarian value overall. Sidgwick's suggestion, then, seems to be that a quality such as a disposition to take revenge could indeed be seen by the utilitarian as a vice,

even though in particular cases it leads to actions of which the utilitarian approves, since the disposition is overall of negative value. It is not entirely clear how Sidgwick sees this as a response to the objection, since this rested on the assumption that revenge may in certain societies be of overall utilitarian value. Perhaps he is saying, somewhat indirectly, that it is not, and that Lecky and others have been misled into thinking that the disposition has value through noticing certain non-standard cases in which acting on it is valuable.

Sidgwick then claims that, 'although, in the view of a Utilitarian, only the useful is praiseworthy, he is not bound to maintain that it is necessarily worthy of praise in proportion as it is useful' because what the utilitarian will consider when deciding whether to praise some quality is not the usefulness of the quality, but the usefulness of the praise (428). Sidgwick's general response to Lecky here is a good one. As he goes on to point out, even if self-love and resentment are generally useful qualities in society, they will exist anyway and do not require the backing of common-sense morality as virtues. In just the same way, he continues, common sense praises humility, even though ignorance of one's qualities is in itself undesirable because of the natural tendency for most people towards arrogance. Here we see the radical nature of the utilitarian project manifesting itself once again. For the hedonistic utilitarian, praise, blame, and other attitudes and practices have value or rationale not in themselves, but only in so far as they promote happiness overall. In this context, Sidgwick's first response to Lecky can be seen as correct, in attempting to provide a naturalistic explanation of these attitudes and practices. For the utilitarian, though there is value in well-being, there really is no such thing as *moral* value; the task, then, is to explain why we think there is. It has to be admitted that as yet utilitarians have largely failed to take on the issue with which Sidgwick struggles in this section.

Sidgwick continues with the same strategy at 430 to offer a resolution to the puzzle that arises within common-sense morality over whether moral excellence consists in finding virtue easy (as Aristotle might claim) or hard (as Kant, or at least a Kantian, might claim) (see 224–5; CD 5.3.2). In the case of kind acts, for example, it is not difficult to see how they might be more productive of happiness if performed without effort by the moral saint; whereas if someone struggles successfully to perform such acts, we recognize that character trait as more generally beneficial, in enabling the moral hero to do her duty in other contexts.

And, Sidgwick points out, a utilitarian will have an answer to the question of whether we should give priority to subjective or to objective rightness in a case in which we have the opportunity to influence someone in the direction of acting against his moral convictions to do what is objectively right. We have to balance the good consequences of the right act against the bad consequences of the moral deterioration of the person's character. Sidgwick's conceptual scheme has here become a little tangled. If by acting against her convictions the person affects her character negatively to the extent that it would have been better never to have done so, it cannot be said that her initial action was objectively right: it would be both objectively and subjectively wrong. But Sidgwick's point is clear enough. In the dilemma in question, the utilitarian has to decide whether or not to encourage the person to act against their conscience purely through comparing the consequences of such encouragement against those of non-encouragement.

In the final section of 4.3, having explained in previous sections how utilitarianism 'controls and completes' the principles of common-sense virtues, Sidgwick expounds further his claim that common sense is unconsciously utilitarian (453–7). He begins by noting again how utilitarianism is standardly used by common sense to resolve conflicts between principles, and disputes as to the scope or definition of any currently accepted principle, referring back to the argument of 3.14 intended to demonstrate the implicit hedonism in the common-sense view of well-being. He then claims that unconscious utilitarianism also explains why different classes of people rank virtues differently, and why different virtues are thought more appropriate for certain classes than for others. Chastity, for example, is thought more important for women, and courage for men, especially if they are soldiers. Further, the variations across time and space between moral codes can be explained by the different effects of actions on happiness, or by the different predictions of those effects or opinions about their significance. One of Sidgwick's examples is taken from Dugald Stewart: the South Sea Islanders do not take theft seriously because little labour is required in their environment for the maintenance of life.

Further, as these predictions become more precise, we can see progress towards a 'perfectly enlightened Utilitarianism', as sympathy with others is extended. There are, as Sidgwick will go on to explain later in the chapter, exceptions. But even these, he suggests, are often only the result of excessive weight's being placed on a sentiment which is indeed useful to some degree,

or the survival of some sentiment useful only in the past. In other words, the moral history of humanity, as Sidgwick sees it, consists in utilitarianism moving from an implicit to an explicit status:

It is therefore not as the mode of regulating conduct with which mankind began, but rather as that to which we can now see that human development has been always tending, as the adult and not the germinal form of Morality, that Utilitarianism may most reasonably claim the acceptance of Common Sense. (456–7)

The fact remains, however, that the dogmatic intuitionist can accept a great deal of Sidgwick's argument about the utility-value of common-sense morality, and even its emergence from 'unconscious utilitarianism'. And she might even agree with Sidgwick that conflicts between non-utilitarian principles, or the scope of such principles, are to be decided with reference to utilitarianism. Sidgwick insists that any principle must be self-contained, of course, so the intuitionist would have to redescribe all her principles as, say, 'justice plus utility', 'veracity plus utility', and so on. What matters is whether, say, the fact that one is breaking a promise, in normal circumstances, counts against one's action. Nothing Sidgwick says about common-sense morality and utility prevents the intuitionist from insisting that it does. In other words, the alleged presence of utilitarianism within common-sense morality, though it might offer various dialectical opportunities for Sidgwick (e.g. the utilitarian principle might be the only common-sense principle to survive the onslaught of book 3), also leaves various hostages to fortune which he failed to recognize by paying insufficient attention to the (huge) difference between rational benevolence as construed by the utilitarian on the one hand, and by the dogmatic intuitionist pluralist on the other.

2.4 The Method of Utilitarianism

Sidgwick is interested not only in the question of whether utilitarianism can be proved. As we have seen, he sees philosophical ethics as importantly practical, and in the case of any particular principle, he will ask what its implications are for making real-life decisions (see CD 1.3–4). According to Sidgwick's dualism of practical reason, as we shall see, there is a stand-off between utilitarianism and egoism. But Sidgwick allows that 'practical reason would still impel us decisively to the performance of duty in the

more ordinary cases in which what is recognised as duty is in harmony with self-interest properly understood' (508). By 'duty' here, Sidgwick means what utilitarianism gives us ultimate reason to do; so even given the dualism, there is a potentially large place left in life for the method recommended by utilitarianism. This both explains and justifies the two substantial chapters—4.4 and 4.5—that Sidgwick dedicates to close discussion of what this method (or these methods) might be.

The obvious utilitarian method is, of course, empirical hedonism (4.4.1). In book 2, Sidgwick carefully teased out the difficulties in applying this method directly in the case of a single individual. But, he suggests, despite the fact that interpersonal comparisons of utility will lead to yet another level of complexity, it might be thought that the arguments he has offered for utilitarianism's constituting a foundation for common-sense morality enable us to see the principles of common-sense morality as 'middle axioms' of the utilitarian method, lying between the higher principle of utilitarianism itself and individual decisions about what to do in particular circumstances (461). (Common-sense principles of prudence will, of course, resolve also many of the practical difficulties arising from direct application of empirical hedonism in the individual case.)

Common-sense morality, however, is not perfectly in alignment with utilitarianism at every point. Having allowed some weight to the psychological explanation of the origin of the moral sentiments in experiences of pleasure and pain (see CD 7.2.1), Sidgwick, with reference to Adam Smith, draws a distinction between sympathy felt with the pain and pleasure of others that results from certain actions, and sympathy with the impulses that lead to action in the first place (463–4). He suggests that moral rules often appear to be compromises between the two kinds of sympathy, which can lead to non-utilitarian results because of the insufficient influence of pleasure and pain on the 'active impulses'. Further, even if our sentiments are entirely the result of accumulated experiences of pleasure and pain, they would lead to a morality consistent with utilitarianism only in so far as they are combined with accurate predictions of the hedonic consequences of particular actions. And in fact it is clear that such predictions have frequently been deeply mistaken because of limited sympathy and knowledge, and distortions resulting from social hierarchies and false religions.

It might be thought that aspects of morality which did not promote survival would fade in the 'struggle for existence' (464–5; see 471–4). But they would represent only one source of inefficiency, and anyway a morality that

effectively preserved a community might still be lacking from the utilitarian point of view. Further, changed circumstances may mean that a morality which is now hedonically the most efficient becomes less so. History shows us that moralities in the past have often been poor means to general happiness, and we should presume that there are similar imperfections in our own morality which our familiarity with it makes it harder for us to see. Indeed, the fact that within our own culture some minority—especially one constituted by ‘persons of enlightenment and special acquaintance with the effects of the conduct judged’—disagrees with common-sense morality should lead us to distrust it in that particular case (466). The morality of common sense is designed for ordinary circumstances, and the question therefore arises of whether special circumstances or individual abilities require a special morality. Further, since many people often see rightness and utilitarian expediency coming apart, we need to consider the cases in which this is thought to be occurring (466–7).

In the end, Sidgwick believes that defining precisely the moral code recommended by utilitarianism will face problems of the same magnitude as he found with the application of empirical hedonism in book 2 (467–71). The nature of human beings themselves varies greatly over time and space, which makes it impossible to construct a universal ideal. If we restrict ourselves to our own time and country, Sidgwick suggests, we face a dilemma: if we take our fellow-citizens as they are, we cannot see them as subjects for a new moral code; while if we abstract away, it is not clear why we should want to design a community for such imaginary beings, unless we assume—quite implausibly—that any code we design will be immediately and appropriately adopted by all. Indeed, working out an ideal may be counter-productive, if attempting to put into practice any new rule serves only to undermine the existing one in the same area. Seeking to approximate to an ideal designed on the assumption that the non-moral aspects of human nature remain unchanged may itself be less productive in its non-responsiveness to the felicitic opportunities presented by change. This presents another dilemma for those constructing an ideal code: we cannot assume human nature to be constant, unless we restrict ourselves to the present or near future; while if we do so restrict ourselves, we must see people’s actual moral sentiments as part of their human nature and no more alterable than non-moral elements of that nature.

It has to be said that, though Sidgwick is of course right about the difficulties of predicting the future, he makes rather heavy weather of these

so-called dilemmas. It is clear that we have to take human beings as they are, while recognizing that human nature includes a capacity to change to some limited degree (as he himself sees: 474; 475–6; 477). But Sidgwick's objections against Spencer's suggestion that utilitarians should aim at describing the rules of some perfect form of society are well taken (470–1): we cannot predict the natures and relations of the individuals in such a society sufficiently precisely, and even if we could, it does not follow that the rules of such a society will be the best rules for us (consider, for example, the need we have for rules governing punishment).⁴⁰ In the second half of the nineteenth century, it was quite common for those influenced by Darwinian theories of evolution to see the end of morality as the preservation of the social organism rather than its happiness (471–4; see Lillehammer 2010). Again, Sidgwick's objections to Stephen's version of this view are right on target: there is no reason to think that preservation maximizes happiness, since many pleasures and pains have nothing to do with preservation; and, given the imperfect state of sociology (see 476), the difficulties in working out the best means to preservation are no less than those in working out the means to maximum happiness. Nor will substituting the idea of 'development' for that of 'preservation' help.

4.4 sets the general background to the utilitarian project of assessment of and gradual improvement to common-sense morality. 4.5 provides further and particular details of that project. Sidgwick's first point is to reassure any reader who might be concerned at the potentially radical implications of utilitarianism that the utilitarian will on the whole respect, conform to, and promote common-sense morality, merely seeking to change the emphasis of that morality towards, for example, greater condemnation of action showing a lack of benevolent concern (475–6; 484).⁴¹ But utilitarians also have a duty to improve that morality wherever they can, and their task cannot be made more tractable by separating off some Millian 'private' sphere which is not to be regulated by moral sanctions, both because there is no such sphere and because, even if there were, an individual, if she is a utilitarian, will want to guide her action within it in the direction of overall happiness (476–9). Nevertheless, we should remember that seeking happiness for oneself and

40. This point became one of the standard objections to so-called 'rule utilitarianism'. See e.g. Hooker 2000b: 80–5.

41. One is reminded here of D. G. Ritchie's remarks on the utilitarianism in Sidgwick's EP (1892: 255): 'The method is Bentham's; but there is none of Bentham's strong critical antagonism to the institutions of his time ... If this is Benthamism, it is Benthamism grown tame and sleek.'

for others on the basis of one's own experience is widely thought to be typical of a sensible person, and even when a person believes in non-hedonic goods she will deal with any conflict between them by reference to empirical hedonism.⁴² Similarly, though the utilitarian will make good use of sciences such as economics or physiology, when it comes to weighing, say, wealth against health, empirical hedonism is our only resource (479).

Here the utilitarian must not forget the possible consequences to her and those close to her of open disapproval of some aspect of common-sense morality (480–4). She needs to be sure that now is the right time to make her case, and that the positive effects on happiness that she envisages will really occur: the new rule may be too complex to be widely adopted, and she must remember that she may be setting a bad example in her attitude to common-sense morality and that going against that morality may have certain destructive effects on her own character. A morality is sustained by general sentiments, and it is much easier to conform to an established social rule than one invented by oneself. Nevertheless, a striking example of consistent utilitarianism can have good effects if the change being recommended is not to the agent's own advantage. Further, when it comes to the mere addition of a new rule to common-sense morality, the utilitarian is acting quite consistently with that morality, since it includes a principle of universal benevolence which in this case the utilitarian is trying to put into practice. She should, however, beware of criticizing others for not adopting the rule, since a less aggressive attitude may be more effective.

Sidgwick then turns to the difficult issue of exceptions to rules, which both utilitarianism and common sense hold to be generally justified. He notes that an exception on general grounds is merely to make the rule itself more precise (485).⁴³ For example, a utilitarian might think that *everyone* should lie in response to an inquiry into how they voted in a secret ballot. But what about cases in which the utilitarian thinks a change to the rules is not justified, while his exceptional action is (486–7)?

One cannot argue that, because an action is one of a set which would be overall harmful, it should not be performed. For taken individually it may be beneficial. Nor can a Kantian object that such an action is forbidden

42. Sidgwick presumably thinks this because he cannot conceive of any other 'common denominator'. He never discusses the question of whether judgement is possible without such a denominator: see CD 1.5 and 6.9.

43. See e.g. Lyons 1965: ch. 3.

because the agent cannot will her maxim of action to be a universal law.⁴⁴ For that the action is not going to be widely imitated may be part of the description of the act in question (see Lipkin 1967).⁴⁵ And common sense itself frequently permits behaviour that would be disastrous if widely imitated, such as celibacy.

Nevertheless, cases in which such deviation was justified would be very rare (487–90). There is a difference between celibacy and, say, lying. In the case of celibacy, my own preference is enough to justify my decision, and we can all recognize celibacy as universally legitimate. But a mere preference cannot justify a lie, and a belief that it could would undermine the moral sentiments. So a utilitarian would have to admit that she is advocating that others do the same as her in similar circumstances, and this could weaken the moral sentiment protecting the principle of veracity in general. In other words, especially given the dangers in moral innovation already mentioned, it will be only very infrequently that a utilitarian can be certain that by approving a violation of a generally useful rule she is not doing more harm than good.

This raises the question of whether the risks of transparency might not justify a utilitarian in acting against a common-sense rule in secret, which Sidgwick answers in a passage which has become somewhat notorious (see e.g. Coady 1994: 132–7; Williams 1995: 165–6):

[O]n Utilitarian principles, it may be right to do and privately recommend, under certain circumstances, what it would not be right to advocate openly; it may be right to teach openly to one set of persons what it would be wrong to teach to others; it may be conceivably right to do, if it can be done with comparative secrecy, what it would be wrong to do in the face of the world; and even, if perfect secrecy can be reasonably expected, what it would be wrong to recommend by private advice or example. (489)

Further, since it is clearly beneficial on the whole for people not to be encouraged to conceal their wrongdoing,

the opinion that secrecy may render an action right which would not otherwise be so should itself be kept comparatively secret; and similarly it seems expedient that the doctrine that esoteric morality is expedient should itself be kept esoteric. Or if this concealment be difficult to maintain, it may be desirable that Common Sense should repudiate the doctrines which it is expedient

44. Sidgwick here refers back to 3.1 [209–10] and 3.13 [385–6].

45. Sidgwick here refers back to 3.7.3 [317–19].

to confine to an enlightened few. And thus a Utilitarian may reasonably desire, on Utilitarian principles, that some of his conclusions should be rejected by mankind generally; or even that the vulgar should keep aloof from his system as a whole, in so far as the inevitable indefiniteness and complexity of its calculations render it likely to lead to bad results in their hands.⁴⁶ (490)

In other words, moral beliefs themselves may be assessed in the light of the principle of utility, and it may maximize happiness if different and conflicting moral opinions are held in the same society at the same time (491–2). Nor need this be seen as entirely inconsistent with common sense. Consider rebellion. On the one hand, common sense recognizes that, to discourage rebellions, there should be a widespread body of opinion that any rebellion should be punished; on the other, it allows that certain rebellions—and hence the beliefs of the rebels—are justified. And common sense may admit that certain people in certain occupations may be permitted to break rules which are held inviolable by society at large. (Sidgwick gives no example, but perhaps has politicians primarily in mind (see PM).) Nevertheless, a utilitarian will recognize the dangers of such conflicts and in most cases seek to resolve them, most often by pointing out that the exception is not what the person intending to break the rule took it to be (since the breaking of such rules is usually done from self-interest rather than benevolence).

At 492–5, Sidgwick returns to the issue of supererogation, understood in terms of the distinction between excellence and duty (see 220–2; CD 5.2.3). Though that notion has no place within utilitarian theory strictly understood, Sidgwick shows that the instrumentalist attitude taken by utilitarianism to praise and blame enables the utilitarian to retain it within common-sense morality. We must peg blame and other negative sanctions to an ‘average standard’, and use the positive sanctions for behaviour above it, always of course seeking to raise the standard as high as possible.⁴⁷ Further, the utilitarian may be required to seek to amend her own society’s moral ideal, and here the risks of so doing are far less than in the case of rules of duty. Moral ideals are usually so vague, and vary so much, that a utilitarian can often claim anyway to be putting an interpretation on an ideal already in existence, though it is true that utilitarians will be more likely than most to seek to advance qualities that cause pleasure in others rather than those that tend on the side of restriction. Indeed, on the whole we can expect that

46. For a defence of Sidgwick’s position, see de Lazari-Radek and Singer 2010; 2014: ch. 10.

47. Sidgwick was fully aware that utilitarianism is significantly more demanding than common-sense morality: see 87.

common sense will approve of any attempt to promote the common good, even if it is somewhat unusual, as long as it is reasonably thought through and not in violation of any perceived moral rule. Such attempts made by utilitarians often take them in the direction of politics, where beneficence can be especially effective, and we see here another contrast with dogmatic intuitionism, according to which virtue can be as admirably expressed in private as in public life.

3 The Dualism of Practical Reason

In his 'Concluding Chapter', Sidgwick draws together the strands of his argument with a view to making a final decision on the relation of the three methods with which he began the *Methods*: egoism, intuitionism, and utilitarianism.⁴⁸ He begins by reminding us of the philosophical intuitionist basis of utilitarianism, and the lack of it for dogmatic intuitionism, noting that the virtues of dogmatic intuitionism can be seen—partly through reflection on the comparative history and origins of morality—as grounded in impartial benevolence or prudence (496–7). The question, then, is that of the relation between egoistic and universalistic hedonism, and the challenge for anyone who wishes to argue for the rationality of morality is to demonstrate a harmony between those two views (497–8).

C. D. Broad, in an influential discussion of the dualism, claims that harmony is unachievable, since the two views are logically inconsistent (Broad 1930: 158–9; 244–5; 253; see Irwin 2009: 528–9). Broad focuses first on 420–1, where Sidgwick claims that an egoist who says that she ought to take her own happiness as her ultimate end is immune to any argument from the utilitarian that her own good can be no more important a part of the total good than that of any other. For she will deny that her own good is good 'from the point of view of the Universe'.

Later, Broad characterizes the inconsistency in terms of different degrees of concern, the utilitarian claiming that I ought to be *equally* concerned about good states of mind wherever they occur, the egoist that I ought to be *more* concerned about a good state in my own mind than a state of the same value in some other mind. This is not an unreasonable view to attribute to

48. For an excellent discussion of the weight Sidgwick placed on systematizing reasons for action, see Richardson 1991.

Sidgwick, but it is one that he should have been reluctant to accept. What matter are the ultimate principles of utilitarianism and egoism, not the decision-procedures or patterns of concern they recommend. Broad's first characterization of the inconsistency, then, is more fundamental:

E: There is no total or universal good. I ought to aim at my good—that is, what is good for me.

U: There is a universal or total good. I ought to aim at this good—that is, what is good overall.

As Broad goes on to point out, no assumption about the world can remove the inconsistency between the two views.⁴⁹ So what is going on in Sidgwick's 'Concluding Chapter'? Why does Sidgwick not suspend judgement on both egoism and utilitarianism, and work further on the question of which is correct?

The answer, I suggest, lies in the metaphor of 'points of view'.⁵⁰ We can remove the contradiction between *E* and *U* as follows:

*E**: From my own point of view, I ought to aim at what is good for me.

*U**: From the point of view of the universe, I ought to aim at what is good overall.

There are, then, two ways of looking at my own good: I can see the distinction between myself and others as fundamental, and accept *E**; or I can see my good as a mere part of the overall good, and accept *U**. The reason that Sidgwick does not suspend judgement on egoism and utilitarianism is that he sees *both* of these perspectives on oneself as ultimately reasonable.⁵¹ Just after the 'distinction' passage, he says:

49. See Frankena 1974: 457–8: 'It seems to me that two ethical principles cannot both be regarded as self-evident if it is in principle possible for them to come into conflict, and that even a postulate of coincidence in practice cannot save them both. For the coincidence might obtain only because of a fortunate accident about the constitution of our world, and not be true of other possible worlds. But this is a hard question and Sidgwick does not consider it.'

50. See Shaver 1999: 81–2; Skorupski 2001: 71. Skorupski might object to my formulations here that they should also include a statement of what *gives* me the reason in question, and that once this is included the two principles will contradict one another (2001: 69). If the egoist claims, for example, that the fact that ϕ -ing promotes my good gives me a reason to ϕ , then the utilitarian will just deny this. But this is to ignore the perspectival context. The egoist is reporting what can be seen from the agent's own self-regarding point of view, while the utilitarian is speaking of what can be seen from the impartial point of view. It is from the third perspective of actual agency that these principles become practical, and indeed they can become practical in a world in which they do not recommend conflicting courses of action.

51. See M3 402, where Sidgwick says the dualism consists in 'the inevitable twofold conception of a human individual as a whole in himself, and a part of a larger whole. There is something that it is reasonable for him to desire, when he considers himself as an independent unit, and

[E]ven if a man admits the self-evidence of the principle of Rational Benevolence, he may still hold that his own happiness is an end which it is irrational for him to sacrifice to any other; and that therefore a harmony between the maxim of Prudence and the maxim of Rational Benevolence must be somehow demonstrated, if morality is to be made completely rational. This latter view, indeed (as I have before said), appears to me, on the whole, the view of Common Sense: and it is that which I myself hold. It thus becomes needful to examine how far and in what way the required demonstration can be effected.⁵²

something again which he must recognize as reasonably to be desired, when he takes the point of view of a larger whole; the former of these objects I call his own ultimate "Good", and the latter Ultimate Good taken universally.' Cited and discussed in Schneewind 1977: 369–70. Note that I am taking 'ultimately' here to mean 'non-derivative', not 'overriding' or 'supreme' (Irwin 2009: 519; see 453–4). Further, these principles of ultimate reasonableness are universal, applying to any possible action by anyone at any time.

52. We do not, then, have to see Sidgwick as holding that any method must be an attempt to state the *only* source of reasons for action. See Roberts 1969: 62; Mackie 1992: 170; Richardson 1991: 196 n. 26; McLeod 2000: 278–9. Brink (1992: 235 n. 8) claims that Sidgwick usually sees a method as an attempt to state 'the right-making characteristic', with a reference to five different pages of the methods. Let me deal with the passages he may have had in mind. At 6, Sidgwick does say that 'the philosopher seeks unity of principle'. But he could allow the combination of egoism and utilitarianism into a single principle. And note that later on 6 he says that 'so far as two methods conflict, one or other of them must be modified or rejected'. This strongly suggests that if there is no conflict, then there is no need to modify or reject either. 83–4 includes the following passage: 'The discussion in the preceding section will have shown that not all the different views that are taken of the ultimate reason for doing what is concluded to be right lead to practically different methods of arriving at this conclusion. Indeed we find that almost any method may be connected with almost any ultimate reason by means of some—often plausible—assumption. Hence arises difficulty in the classification and comparison of ethical systems; since they often appear to have different affinities according as we consider Method or Ultimate Reason. In my treatment of the subject, difference of Method is taken as the paramount consideration: and it is on this account that I have treated the view in which Perfection is taken to be the Ultimate End as a variety of the Intuitionism which determines right conduct by reference to axioms of duty intuitively known; while I have made as marked a separation as possible between Epicureanism or Egoistic Hedonism, and the Universalistic or Benthamite Hedonism to which I propose to restrict the term Utilitarianism.' This passage does raise the question of why Sidgwick puts such weight on methods rather than principles (see CD 1.4). But nothing here implies that there is only one ultimate principle. I can find no other passage on 84 that implies this. At 174, Sidgwick says that complete 'abdication of self-love' is not possible for a sane person 'who still regards his own interest as *the* reasonable ultimate end of his actions' (my italics). But he is speaking here of the egoist, who of course holds that there is only one such end. At 403, Sidgwick is seeking to explain the aversion of common sense to the idea that happiness alone is our 'sole ultimate end and standard of right conduct' (402). He notes that people usually take this idea to be that one should seek one's own individual happiness at the expense, if necessary, of that of others, but goes on to claim that this 'supreme aim' is in various ways unsatisfactory. Sidgwick's point here is against egoism, but also against a dualistic view that combines egoism with universal benevolence. There is once again no implication that ethical theories must state a single right-making characteristic. Further, since egoism and utilitarianism might anyway be combined into a single principle, the characteristic of an action that it promotes one's own happiness and general happiness can be seen as single. (Brink also refers to two passages which he

So this is why it becomes so important to demonstrate the practical consistency or ‘harmony’ of the two perspectives. If each perspective were correct, then I always have reason to promote my own good (and not doing so would be unreasonable) and I always have reason to promote the good impartially (and not doing so would be unreasonable). For me to be reasonable, I must⁵³ always promote both my own good and the overall good: I have ultimate reason to do both.⁵⁴

sees as potential counter-evidence to his interpretation. But I can see no such evidence in 77 n. At 421, Sidgwick does indeed draw a distinction between the claim that rational benevolence is one principle among others, and the utilitarian claim that it is the ‘sole or supreme’ principle. But this distinction seems consistent both with Brink’s ‘single right-making characteristic’ interpretation and its denial.)

53. I am avoiding any significant contrast here between what is reasonable and what I must do; cf. Phillips 2011: 118. Phillips uses the notion of reasonableness to construct a ‘permissive’ interpretation of the dualism, according to which it is rationally permissible to promote one’s own good and rationally permissible to promote the overall good. Phillips admits that the textual evidence is against the permissive interpretation, but suggests that we should adopt it on the ground that Sidgwick’s own arguments for egoism and utilitarianism ‘exclusively interpreted’ fail and provide stronger support for a qualified version of the permissive view. As Phillips himself says, a proponent of the ‘standard view’, stated in terms of ‘must’, ‘can respond that, even if this is what Sidgwick *ought* to think, it is not what he (on the whole) does think’ (2011: 140). I would add that I think there is less evidence than Phillips claims for a permissive interpretation (Crisp 2013c: paras 27–8). Phillips cites a passage from the first edition of the *Methods* which expresses the axiom of justice with reference to what is ‘right’, ‘reasonable’, ‘the dictate of reason’, and ‘my duty’ (MI 470, cited at 116), and notes that ‘reasonable’ in ordinary language tends to suggest permission’. But if Sidgwick were here mixing concepts of requirement with a concept of permission, he would of course be deeply confused. ‘Reasonable’ for him is, in such contexts, a technical term: the reasonable action is the one there is strongest ultimate reason to do (see e.g. 78 for elucidation of reasonableness in terms of ultimate reasons). Nor does it make a difference that he uses ‘reasonable’ elsewhere in the ordinary sense (Phillips 2011).

My own view (see CD 1.3) is that Sidgwick tends to use too many concepts to express his position, the passage here quoted by Phillips being an excellent example of that. That position could be stated purely in terms of what we have ultimate reason to do, with no reference to permission, requirement, ‘ought’, duty, or whatever. So—*pace* Phillips (2011: 153 n. 17)—I would not want to attribute a ‘permissive’ interpretation to Sidgwick, nor even to advocate myself such a version of a ‘dual source’ view of practical reasons. For me to know what to do, all that is required is that I know what I have strongest ultimate reason to do. To ask whether I am permitted or required to act in that way is to ask an unnecessary and potentially confusing question.

54. Roberts 1969: 62; Schneewind 1977: 373; Mackie 1992: 170; Frankena 1992: 193–4. Also Richardson 1991: 196 n. 26; McLeod 2000: 280. Parfit (2011: 1.131) understands Sidgwick’s dualism constructively: one always has reason to do what is best impartially, unless some other act would be best for one, in which case one has sufficient reason to act in either way; see also Baier 1991: 202. According to Parfit (2011.1.131), Sidgwick held that impartial and self-interested reasons are not comparable, and so, when these reasons conflict, we have sufficient reason to act in either way. I agree that these reasons are incomparable in that they cannot be weighed against one another. This is a consequence of the universal application of the relevant principles. But Sidgwick clearly believed that any conflict led not to sufficiency of reason to act in either way, but to practical chaos: ‘so far as two methods conflict, one or other of them must be modified or rejected’ (6; see 12; Seth 1901: 176; McTaggart 1906: 413; Irwin 2009: 521; Nakano-Okuno 2011: 158).

Sidgwick implicitly recognizes a ‘third perspective’ of practical reason: that which takes the deliverances of both points of view into account, leading to an intention to act.⁵⁵ The notion of a point of view here, then, is epistemological. It is not, for example, as if I have a reason to promote my own happiness only when I am attending to the distinction between individuals. Why did Sidgwick not consider a version of the dualism which, in cases of conflict, weighs the strength of each reason against the other (see Phillips 2011: 132–3)? So, for example, I might conclude in one case that I have strongest reason to promote the overall good at some small cost to myself, and in another that I have strongest reason to promote my own good to a large degree at some small cost to the overall good. This position, according to which we have both egoistic and utilitarian *pro tanto* reasons for action, strikes me as very plausible (Crisp 2006: ch. 5; see Richardson 1991: 184; McLeod 2000: 287; Irwin 2009: 528; Phillips 2011: 140–51). My guess is that Sidgwick would have rejected it as a version of aesthetic intuitionism, since it relies on practical judgement at the meta-level to weigh egoistic and utilitarian reasons against one another. So the antipathy to such judgement that led Sidgwick to miss the strengths of a pluralistic and reflective moral view such as that later developed by Ross also caused him to miss a way to avoid practical reason’s becoming chaotic.

Because complete coincidence between egoism and utilitarianism is required for the harmony Sidgwick sought, it is not enough to point out that both E^* and U^* recommend *general* adherence to the rules of common-sense morality (498–9). Sidgwick goes on to discuss the claim by some utilitarians that this coincidence is ensured by the priority of sympathy as a component of human happiness (499–503). Sidgwick refers to Mill in a note (499–500 n), and reiterates his distinction between sympathy’s role in producing pleasures and pains and its role in causing an impulse to action. As he points out, for sympathy to guarantee the coincidence of E^* and U^* , it must not merely motivate altruistic action but provide maximal happiness for the agent.

Sidgwick recognizes that sympathy can be a source of happiness, and that such sympathy tends to play a role, in the mind of a utilitarian, in the ‘moral feelings’ that concern social conduct. This enables the utilitarian to avoid the objection (often made against Kantian theories in particular) that her theory requires her to sacrifice herself to an ‘impersonal law’ rather than for

55. See Frankena 1974: 456, which refers to M7 404 and 404 n; also 1992: 195. For the opposite view, see Parfit 2011: I.134.

others she cares about. He claims also that most people's happiness would in fact be promoted were they to cultivate a greater degree of sympathy in themselves:

[T]he selfish man misses the sense of elevation and enlargement given by wide interests; he misses the more secure and serene satisfaction that attends continually on activities directed towards ends more stable in prospect than an individual's happiness can be; he misses the peculiar rich sweetness, depending upon a sort of complex reverberation of sympathy, which is always found in services rendered to those whom we love and who are grateful. He is made to feel in a thousand various ways, according to the degree of refinement which his nature has attained, the discord between the rhythms of his own life and of that larger life of which his own is but an insignificant fraction.⁵⁶ (501)

But even this is insufficient to provide the complete coincidence required between E^* and U^* (501–2), since the claims made about conscience (170–5) carry across to sympathy. A sacrifice of one's life for the general good, for example, could not plausibly be said to advance one's happiness, and the fact that our most intense sympathy is for those close to us increases the motivational opposition to impartial utilitarian duty (nor should we think that attempts to increase the impartiality of our sympathy would themselves be recommended by utilitarianism). The same is true in less unusual cases. Alleviating the suffering of others, for example, will be required by utilitarianism, but sympathy here will be if anything a source of pain rather than pleasure to the agent, and, though it may be counterbalanced by the pleasures of benevolence and so on, an alternative life would often be hedonically more valuable for the agent.

Sidgwick then moves to another argument put forward by the utilitarians of his day: that utilitarianism is the law of God, to be enforced through a system of divine reward and punishment that will underpin the practical consistency of E^* and U^* (503–6).⁵⁷ This raises the question of what justifies such beliefs. Sidgwick sees the issue of revelation as beyond his remit, though he cannot resist pointing out that most arguments from revelation have

56. Characteristically, Sidgwick qualifies this passage in a note (501 n): 'Some few thoroughly selfish persons appear at least to be happier than most of the unselfish; and there are other exceptional natures whose chief happiness seems to be derived from activity, disinterested indeed, but directed toward other ends than human happiness.'

57. Frankena (1956: 543–4; 1974: 459) claims, on the basis of the less pessimistic conclusions of the later editions, that Sidgwick was ready to espouse this as a solution to the problem of the dualism, if hesitantly. But the weight Sidgwick places on self-evidence suggests that he remains pessimistic, though prepared to entertain the coherentist response to scepticism.

been to non-utilitarian conclusions. But whether reason itself can give us knowledge of God is a matter of relevance to philosophical ethics as well as theology. Sidgwick rejects the view that moral rules should be seen merely as the commands of a divine lawgiver on the ground that God himself is understood as a moral agent bound by the rules, though Sidgwick is prepared to entertain the position that through intuition we may learn that God commands us to obey certain moral rules independent of his commands or to pursue the same end as he himself pursues—that is, universal happiness.

At this point the problem of evil arises. If God is good and omnipotent, and morality is utilitarian, why is there so much suffering in the world? Whether it is because of certain limitations which do not detract from God's omnipotence as ordinarily understood is a question Sidgwick again refers to theology. But he is concerned to point out that it is not only pleasure which is mingled with its opposite in the world as it is; perfection is balanced by imperfection, virtue by vice. So no special objection to hedonism can be mounted in natural theology.

So can belief in the existence of such a God be justified on purely ethical grounds alone (506–9)? Sidgwick has to confess, with some regret, that he does not see it as self-evident that performance of duty will be rewarded and violation punished, whether by God or in any other way. Thus he feels forced

to admit an ultimate and fundamental contradiction in our apparent intuitions of what is Reasonable in conduct; and from this admission it would seem to follow that the apparently intuitive operation of the Practical Reason, manifested in these contradictory judgments, is after all illusory. (508)

In other words, if egoism and utilitarianism, when construed in the light of the facts, contradict one another, neither of them can in the end be said to be self-evident.⁵⁸ This explains the pessimism of the famous final sentence of the first edition: 'the Cosmos of Duty is thus really reduced to a Chaos: and the prolonged effort of the human intellect to frame a perfect ideal of rational conduct is seen to have been fore-doomed to inevitable failure' (MI 473).

Later editions were less pessimistic.⁵⁹ Having made the claim above, Sidgwick swiftly notes that he is not to be taken as suggesting that 'it would

58. See Seth 1901: 180; Frankena 1974: 457–8; 1992: 194; Singer, M. 1974: 446; Nakano-Okuno 2011: 158.

59. Note what Sidgwick said in a frank letter of 1880 to an old school friend: '[I]f I am asked whether I believe in a God, I should really have to say that I do not know—that is, I do not know whether I *believe* or merely *hope* that there is a moral order in this universe that we know,

become reasonable for us to abandon morality altogether' (508). But, it has to be pointed out, it is hard to see how he is not suggesting that it would not be unreasonable to do so.⁶⁰ Sidgwick himself admits that, though we might still be prompted to do our duty on the basis of self-interest and sympathy, any conflict between self-interest and duty would have to be decided (in practice) by the weight of the 'non-rational impulses' in play.

In the final paragraph of the *Methods*, Sidgwick raises the question of whether the very fact that some hypothesis is required to avoid a contradiction in an important area of thought is itself reason for accepting that hypothesis. Once again, however, he refers the issue elsewhere, this time to 'general philosophy'. Despite the fact that the outcome of his ethical project depends fundamentally on this issue, there is no extended discussion of it in his other works.⁶¹ This is especially odd given his readiness to offer coherentist elements a place in his epistemology in the *Methods*. Sidgwick was leaving the development of his project to posterity, and the words of his friend F. W. H. Myers, written shortly after Sidgwick's death, seem especially apposite: '[H]e pointed to a definite spot; he vigorously drove in the spade; he upturned a shining handful; and he left us as his testament, *Dig here*.'⁶²

a supreme principle of Wisdom and Benevolence, guiding all things to good ends, and to the happiness of the good. I certainly *hope* that this is so, but I do not think it capable of being *proved*. All I can say is that no opposed explanation of the origin of the cosmos—for instance, the atomistic explanation—seems to me even plausible, and that I cannot accept life on any other terms, or construct a rational system of my own conduct except on the basis of this faith. 'You will say, perhaps, "the question is not whether we should *like*, or find it *convenient* to believe in a God, but whether such belief is *true*." To this I answer, "What criterion have you of the truth of any of the fundamental beliefs of science, except that they are consistent, harmonious with other beliefs that we find ourselves naturally impelled to hold." And this is precisely the relation that I find to exist between Theism and the whole system of my moral beliefs. Duty is to me as real a thing as the physical world, though it is not apprehended in the same way; but all my apparent knowledge of duty falls into chaos if my belief in the moral government of the world is conceived to be withdrawn.

'Well, I cannot resign myself to disbelief in duty; in fact, if I did, I should feel that the last barrier between me and complete philosophical scepticism, or disbelief in truth altogether, was broken down. Therefore I sometimes say to myself "I believe in God"; while sometimes again I can say no more than "I hope this belief is true, and I must and will act as if it was"' (Sidgwick and Sidgwick 1906: 347–8). See also Schneewind 1977: 374–9.

60. At 200 n. 2, Sidgwick appears to agree with Butler that someone under two conflicting obligations is in fact under no obligation.

61. Schneewind 1977: 378. The issue is briefly mentioned again in NET 605; see also CD chapter 7 n. 59. Sidgwick did of course engage in serious parapsychological research as a potential source of evidence for the existence of God.

62. Myers 1904: 108; cited in Schultz 2004: 719.

Bibliography

- Acocella, G. 2007. 'Divine Justice and Ethical Experience: Reflection on a Page Written by Sidgwick', in Bucolo, Crisp, and Schultz (eds), 40–6.
- Acocella, G. 2011. 'Ethics and Politics in Two Essays by Henry Sidgwick', in Bucolo, Crisp, and Schultz (eds), 38–54.
- Adams, R. 1985. 'Involuntary Sins', *Philosophical Review* 94: 3–31.
- Albee, E. 1901. 'An Examination of Professor Sidgwick's Proof of Utilitarianism', *Philosophical Review* 10: 251–60.
- Albee, E. 1902. *A History of English Utilitarianism*. London: Swan Sonnenschein, chs 16–18.
- Aquinas, T. 1947. *Summa Theologica*, trans. Fathers of the Dominican Province, 3 vols. New York: Benziger Bros.
- Aristotle 1894. *Ethica Nicomachea*, ed. I. Bywater. Oxford: Clarendon Press.
- Aristotle 1958. *Topica et Sophistici Elenchi*, ed. W. D. Ross. Oxford: Clarendon Press.
- Baier, K. 1991. 'Egoism', in P. Singer (ed.), *A Companion to Ethics*. Oxford: Blackwell, 197–204.
- Bailey, O. 2010. 'What Knowledge is Necessary for Virtue?', *Journal of Ethics and Social Philosophy* 4: 1–18.
- Bain, A. 1868. *Mental and Moral Science: A Compendium of Psychology and Ethics*, 2nd edn. London: Longmans & Green.
- Bain, A. 1876. 'Mr Sidgwick's Methods of Ethics', *Mind* O.S. 1: 179–97.
- Baker, L. R. 2013. *Naturalism and the First-Person Perspective*. New York: Oxford University Press.
- Barba, G. 2011. 'Luxury and Happiness in Henry Sidgwick', in Bucolo, Crisp, and Schultz (eds), 56–92.
- Barbour, G. F. 1908. 'Green and Sidgwick on the Community of the Good', *Philosophical Review* 17: 149–66.
- Barker, H. 1902. 'A Recent Criticism of Sidgwick's *The Methods of Ethics*', *Philosophical Review* 11: 607–14.
- Barnes, J. 1976. 'Introduction', in Aristotle, *Ethics*, trans. J. Thomson, rev. H. Tredennick. Harmondsworth: Penguin, 9–43.
- Barratt, A. 1877a. 'The "Suppression" of Egoism', *Mind* O.S. 2: 167–86.
- Barratt, A. 1877b. 'Ethics and Politics', *Mind* O.S. 2: 453–76.
- Barry, B. 1989. *Theories of Justice*. London: Harvester-Wheatsheaf.
- Blanshard, B. 1974. 'Sidgwick the Man', *Monist* 58: 349–70.

- Bok, S. 2000. 'Henry Sidgwick's Practical Ethics', *Utilitas* 12: 361–78.
- Bradley, F. H. 1935. 'Mr Sidgwick's Hedonism', London: H. S. King (1877); repr. in his *Collected Essays*, 2 vols. Oxford: Clarendon Press, 71–128.
- Brink, D. 1988. 'Sidgwick's Dualism of Practical Reason', *Australasian Journal of Philosophy* 66: 291–307.
- Brink, D. 1992. 'Sidgwick and the Rationale for Rational Egoism', in Schultz (ed.), 199–240.
- Brink, D. 1994. 'Common Sense and First Principles in Sidgwick's *Methods*', *Social Philosophy and Policy* 11: 179–201.
- Broad, C. D. 1930. *Five Types of Ethical Theory*. London: Kegan Paul, Trench, Trübner.
- Broad, C. D. 1938a. 'Henry Sidgwick', *Hibbert Journal* 37: 25–43.
- Broad, C. D. 1938b. 'Henry Sidgwick and Psychical Research', *Proceedings of the Society for Psychical Research* 45: 131–61.
- Broome, J. 2007. 'Wide or Narrow Scope?', *Mind* 116: 359–70.
- Bucolo, P. 2007. 'Theism and Sanctions', in Bucolo, Crisp, and Schultz (eds), 50–100.
- Bucolo, P. 2011. 'Theism as a Model of Reconciliation', in Bucolo, Crisp, and Schultz (eds), 94–144.
- Bucolo, P., R. Crisp, and B. Schultz (eds) 2007. *Henry Sidgwick: Happiness and Religion*. Catania: Dipartimento di Scienze Umane dell'Università di Catania.
- Bucolo, P., R. Crisp, and B. Schultz (eds) 2011. *Henry Sidgwick: Ethics, Psychics, Politics*. Catania: C.U.E.C.M.
- Butler, J. 1865. *Three Sermons on Human Nature: Dissertation on Virtue*, ed. W. Whewell, 4th edn. London: Longmans.
- Calderwood, H. 1876. 'Mr Sidgwick on Intuitionism', *Mind* O.S. 2: 197–206.
- Calderwood, H. 1896. 'The Relation of Intuitionism to the Ethical Doctrine of Self-realization', *Philosophical Review* 5: 337–51.
- Carlson, G. 1988. 'Parfit, Sidgwick, and Divided Reason', *Philosophia* 18: 247–52.
- Christiano, T. 1992. 'Sidgwick on Desire, Pleasure, and the Good', in Schultz (ed.), 261–78.
- Coady, A. J. 1994. 'Henry Sidgwick', in C. Ten (ed.), *The Nineteenth Century*, Routledge History of Philosophy, vol. 7. London: Routledge, 122–47.
- Cohen, G. A. 1995. *Self-ownership, Freedom, and Equality*. New York: Cambridge University Press.
- Collini, S. 1992. 'The Ordinary Experience of Civilized Life: Sidgwick's Politics and the Method of Reflective Analysis', in Schultz (ed.), 333–67.
- Collini, S. 2001. 'My Roles and their Duties: Sidgwick as Philosopher, Professor and Public Moralist', in Harrison (ed.), 9–49.
- Collini, S. 2004. 'Henry Sidgwick (1838–1900)', *Dictionary of National Biography*. Oxford: Oxford University Press.
- Cremaschi, S. 2006. 'Sidgwick e il Progetto di un'Etica Scientifica', *Etica & Politica* 8: 1–36.
- Cremaschi, S. 2008. "'Nothing to invite or to reward a separate examination": Sidgwick and Whewell', *Etica & Politica* 10: 137–84.

- Cremaschi, S. 2011. 'As Boys Pursue the Rainbow: Whewell's Independent Morality versus Sidgwick's Dogmatic Intuitionism', in Bucolo, Crisp, and Schultz (eds), 146–234.
- Crisp, R. 1990. 'Sidgwick and Self-interest', *Utilitas* 2: 267–80.
- Crisp, R. 1994. 'Aristotle's Inclusivism', *Oxford Studies in Ancient Philosophy* 12: 111–36.
- Crisp, R. 1997. *Mill on Utilitarianism*. London: Routledge.
- Crisp, R. 2000. 'Review of J. Kupperman, *Value ...and What Follows*', *Philosophy* 75: 458–62.
- Crisp, R. 2001. 'Sanctions in Bentham, Mill, and Sidgwick', in Harrison (ed.), 117–22.
- Crisp, R. 2002. 'Sidgwick and the Boundaries of Intuitionism', in P. Stratton-Lake (ed.), *Ethical Intuitionism: Re-Evaluations*. Oxford: Clarendon Press, 56–75.
- Crisp, R. 2006. *Reasons and the Good*. Oxford: Clarendon Press.
- Crisp, R. 2007a. 'Intuitionism and Disagreement', in M. Timmons, J. Greco, and A. Mele (eds), *Rationality and the Good: Critical Essays on the Ethics and Epistemology of Robert Audi*. Oxford: Oxford University Press, 31–9.
- Crisp, R. 2007b. 'Sidgwick's Hedonism', in Bucolo, Crisp, and Schultz (eds), 104–56.
- Crisp, R. 2011a. 'Pleasure and Hedonism in Sidgwick', in T. Hurka (ed.), *Underivative Duty*. New York: Oxford University Press, 26–44.
- Crisp, R. 2011b. 'Reasonable Disagreement: Sidgwick's Principle and Audi's Intuitionism', in J. G. Hernandez (ed.), *The New Intuitionism*. London: Continuum, 151–68.
- Crisp, R. 2012a. 'Naturalism: Feel the Width', in S. Nuccetelli and G. Seay (eds), *Naturalism and Non-naturalism in Ethics: Current Debates*. Cambridge: Cambridge University Press, 58–69.
- Crisp, R. 2012b. Response on PEASoup blog to de Lazari-Radek and Singer 2012. <<http://peasoup.typepad.com/peasoup/2012/12/ethics-discussions-at-peasoup-katarzyna-de-lazari-radek-and-peter-singer-the-objectivity-of-ethics-1.html>>.
- Crisp, R. 2013a. 'Supererogation and Virtue', *Oxford Studies in Normative Ethics* 3: 13–34.
- Crisp, R. 2013b. 'Homeric Ethics', in R. Crisp (ed.), *Oxford Handbook of the History of Ethics*, Oxford: Oxford University Press, 1–20.
- Crisp, R. 2013c. 'Metaphysics, Epistemology, Utilitarianism, Intuitionism, and Egoism: A Response to Phillips on Sidgwick', *Revue d'études benthamiennes* [online] 12. <<http://etudes-benthamiennes.revues.org/671>>.
- Crisp, R. 2014. 'Taking Stock of Utilitarianism', *Utilitas* 26: 231–49.
- Crisp, R. 2015a. 'A Third Method of Ethics?', *Philosophy and Phenomenological Research* 90: 257–73.
- Crisp, R. 2015b. 'Keeping Things Simple', in I. Hirose and A. Reisner (eds), *Weighing and Reasoning: A Festschrift for John Broome*. Oxford: Clarendon Press, 140–55.

- Crisp, R. forthcoming. 'Prudential and Moral Reasons', in D. Star (ed.), *Oxford Handbook of Reasons and Normativity*. Oxford: Oxford University Press.
- Crisp, R. and M. Slote (eds) 1997. *Oxford Readings on Virtue Ethics*. Oxford: Oxford University Press.
- Crossley, D. 2002. 'Henry Sidgwick', in P. Dematteis, P. Fosl, and L. McHenry (eds), *British Philosophers, 1800–2000*. Farmington, MI: Thomson Gale.
- Csikszentmihalyi, M. 1991. *Flow: The Psychology of Optimal Experience*. New York: Harper Perennial.
- Curtis, D. 1986. 'A Class and State Analysis of Henry Sidgwick's Utilitarianism', *Philosophy and Social Criticism* 11: 259–96.
- Danto, A. 1963. 'What We Can Do', *Journal of Philosophy* 60: 435–45.
- Darwall, S. 1974. 'Pleasure as Ultimate Good in Sidgwick's Ethics', *Monist* 58: 475–89.
- Darwall, S. 2000. 'Sidgwick, Concern, and the Good', *Utilitas* 12: 291–306.
- Daurio, J. 1997. 'Sidgwick on Moral Theories and Common Sense Morality', *History of Philosophy Quarterly* 14: 425–45.
- Davidson, D. 1963. 'Actions, Reasons, and Causes', *Journal of Philosophy* 60: 685–700.
- Deigh, J. 1992. 'Sidgwick on Ethical Judgement', in Schultz (ed.), 241–58.
- Deigh, J. 2004. 'Sidgwick's Conception of Ethics', *Utilitas* 16: 168–83.
- Deigh, J. 2007. 'Sidgwick's Epistemology', *Utilitas* 19: 435–46.
- Deigh, J. 2010. 'Some Further Thoughts on Sidgwick's Epistemology', *Utilitas* 22: 78–89.
- de Lazari-Radek, K. and P. Singer 2010. 'Secrecy in Consequentialism: A Defence of Esoteric Morality', *Ratio* 23: 34–58.
- de Lazari-Radek, K. and P. Singer 2012. 'The Objectivity of Ethics and the Unity of Practical Reason', *Ethics* 123: 9–31.
- de Lazari-Radek, K. and P. Singer 2014. *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. Oxford: Oxford University Press.
- Di Nuovo, S. 2011. 'Starting with Sidgwick: Reflecting on the Phenomenal', in Bucolo, Crisp, and Schultz (eds), 258–86.
- Donagan, A. 1992. 'Sidgwick and Whewellian Intuitionism: Some Enigmas', repr. in Schultz (ed.), 123–42.
- Dreier, J. 2004. 'Why Ethical Satisficing Makes Sense and Rational Satisficing Doesn't', in M. Byron (ed.), *Satisficing and Maximizing: Moral Theorists on Practical Reason*. Cambridge: Cambridge University Press, 131–54.
- Durand, K. 2002. *Sidgwick's Utility and Whitehead's Virtue: Metaphysics and Morality*. Lanham: University Press of America.
- Dworkin, R. 1981. 'What is Equality? Part 1: Equality of Welfare', *Philosophy & Public Affairs* 10: 185–246.
- Edgeworth, F. 1877. *New and Old Methods of Ethics*. Oxford: Parker.
- Ezorsky, G. 1974. 'Unconscious Utilitarianism', *Monist* 58: 468–74.
- Fowler, T. 1885. 'Professor Sidgwick on "Progressive Morality"', *Mind* O.S. 10: 481–6.

- Fowler, T. 1887. *The Principles of Morals*, pt. 2. London: Macmillan.
- Fowler, T. and L. Selby-Bigge 1890. 'Some Fundamental Ethical Controversies', *Mind* O.S. 15: 89–99.
- Frankena, W. 1956. 'Henry Sidgwick', in V. T. A. Ferm (ed.), *Encyclopedia of Morals*. New York: Philosophical Library, 539–44.
- Frankena, W. 1974. 'Sidgwick and the Dualism of Practical Reason', *Monist* 58: 449–67.
- Frankena, W. 1992. 'Sidgwick and the History of Ethical Dualism', in Schultz (ed.), 174–98.
- Frankena, W. 2000. 'The Methods of Ethics, Edition 7, Page 92, Note 1', *Utilitas* 12: 278–91.
- Frankfurt, H. 1969. 'Alternate Possibilities and Moral Responsibility', *Journal of Philosophy* 66: 829–39.
- Frey, R. 1977. 'Act-Utilitarianism: Sidgwick or Bentham and Smart?', *Mind* 86: 95–100.
- Garcia, J. 1987. 'Why Sidgwick's Project Had to Fail', *History of Philosophy Quarterly* 4: 79–91.
- Gauld, A. 2007. 'Henry Sidgwick, Theism, and Psychical Research', in Bucolo, Crisp, and Schultz (eds), 160–256.
- Geninet, H. 2011. 'Henry Sidgwick and Charles Richet', in Bucolo, Crisp, and Schultz (eds), 288–322.
- Giarrizzo, G. 2011. 'A Desirable Democracy', in Bucolo, Crisp, and Schultz (eds), 324–44.
- Gibbard, A. 1982. 'Inchoately Utilitarian Common Sense: The Bearing of a Thesis of Sidgwick's on Moral Theory', in H. B. Miller and W. H. Williams (eds), *The Limits of Utilitarianism*. Minneapolis: University of Minnesota Press, 71–85.
- Gomberg, P. 1986. 'Self and Others in Bentham and Sidgwick', *History of Philosophy Quarterly* 3: 437–48.
- Gray, J. 1984. 'Indirect Utility and Fundamental Rights', *Social Philosophy and Policy* 1: 73–91.
- Green, T. H. 1877. 'Hedonism and Ultimate Good', *Mind* O.S. 2: 266–9.
- Gurney, E. 1887. *Tertium Quid: Chapters on Various Disputed Questions*, 2 vols. London: Kegan Paul, Trench.
- Gyzicki, G. von. 1890. 'Review of *Methods of Ethics*', *International Journal of Ethics* 1: 120–1.
- Haack, S. 1993. *Evidence and Inquiry: Toward Reconstruction in Epistemology*. Oxford: Blackwell.
- Hardin, C. L. 1983. 'Colors, Normal Observers and Standard Conditions', *Journal of Philosophy* 80: 806–13.
- Hardin, R. 1992. 'Common Sense at the Foundations', in Schultz (ed.), 143–60.
- Harrison, R. (ed.) 1996. 'Cambridge Philosophers VI: Henry Sidgwick', *Philosophy* 71: 423–38.
- Harrison, R. (ed.) 2001a. *Henry Sidgwick*. Oxford: Oxford University Press.
- Harrison, R. 2001b. 'The Sanctions of Utilitarianism', in Harrison (ed.), 93–116.

- Harsanyi, J. 1982. 'Morality and the Theory of Rational Behaviour', in A. Sen and B. Williams (ed.), *Utilitarianism and Beyond*. Cambridge: Cambridge University Press, 39–62.
- Havard, W. 1959. *Henry Sidgwick and Later Utilitarian Political Philosophy*. Gainesville: University of Florida Press.
- Hayward, F. H. 1901a. *The Ethical Philosophy of Sidgwick: Nine Essays, Critical and Expository*. London: Swan Sonnenschein.
- Hayward, F. H. 1901b. 'The True Significance of Sidgwick's "Ethics"', *International Journal of Ethics* 11: 175–87.
- Hayward, F. H. 1901c. 'A Reply [to E. E. Constance Jones]', *International Journal of Ethics* 11: 360–5.
- Hills, A. 2010. *The Beloved Self: Morality and the Challenge from Egoism*. Oxford: Oxford University Press.
- Hobbes, T. 2012. *Leviathan*, ed. N. Malcolm, vol. 2. Oxford: Clarendon Press.
- Holley, D. 2002. 'Sidgwick's Problem', *Ethical Theory and Moral Practice* 5: 45–65.
- Holtug, N. 2001. 'On the Value of Coming into Existence', *Journal of Ethics* 5: 361–84.
- Hooker, B. 2000a. 'Sidgwick and Common-sense Morality', *Utilitas* 12: 347–60.
- Hooker, B. 2000b. *Ideal Code, Real World*. Oxford: Clarendon Press.
- Hume, D. 1998. *An Enquiry Concerning the Principles of Morals*, ed. T. Beauchamp. New York: Oxford University Press.
- Hume, D. 2007. *A Treatise of Human Nature*, ed. D. F. Norton and M. F. Norton, vol. 1. Oxford: Clarendon Press.
- Hurka, T. 2001. *Virtue, Vice, and Value*. Oxford: Oxford University Press.
- Hurka, T. 2003. 'Moore in the Middle', *Ethics* 113: 599–628.
- Hurka, T. 2014. 'Sidgwick on Consequentialism and Deontology: A Critique', *Utilitas* 26: 129–52.
- Hurley, P. 2009. *Beyond Consequentialism*. Oxford: Oxford University Press.
- Irwin, T. 1992. 'Eminent Victorians and Greek Ethics: Sidgwick, Green, and Aristotle', in Schultz (ed.), 279–310.
- Irwin, T. 2007. *The Development of Ethics*, vol. 1: *From Socrates to the Reformation*. Oxford: Oxford University Press.
- Irwin, T. 2009. *The Development of Ethics*, vol. 3: *From Kant to Rawls*. Oxford: Oxford University Press.
- Jackson, F. and R. Pargetter 1986. 'Oughts, Options, and Actualism', *Philosophical Review* 95: 233–55.
- James, D. 1970. *Henry Sidgwick: Science and Faith in Victorian England*. London: Oxford University Press.
- Jones, E. E. Constance 1901. 'Mr Hayward's Evaluation of Professor Sidgwick's Ethics', *International Journal of Ethics* 11: 354–60.
- Jones, E. E. Constance 1904. 'Professor Sidgwick's Ethics', *Proceedings of the Aristotelian Society* N.S. 4: 32–52.
- Jones, E. E. Constance 1909. 'Professor Henry Sidgwick's Philosophical Intuitionism', *Bericht über den III internationalen Kongress für Philosophie*, ed. T. Elsenhans. Heidelberg: Carl Winter, 920–7.

- Jones, E. E. Constance 1918. 'Practical Dualism', *Proceedings of the Aristotelian Society* N.S. 18: 317–28.
- Kagan, S. 1989. *The Limits of Morality*. Oxford: Clarendon Press.
- Kahneman, D. 1999. 'Objective Happiness', in Kahneman, Diener, and Schwarz (eds), 3–25.
- Kahneman, D., E. Diener, and N. Schwarz 1999a. 'Preface' to Kahneman, Diener, and Schwarz (eds), ix–xii.
- Kahneman, D., E. Diener, and N. Schwarz (eds) 1999b. *Well-Being: The Foundations of Hedonic Psychology*. New York: Russell Sage Foundation.
- Kahneman, D. 2000. 'Experienced Utility and Objective Happiness: A Moment-based Approach', in A. Tversky and D. Kahneman (eds), *Choices, Values, and Frames*. New York: Cambridge University Press, 673–92.
- Kant, I. 1871. *The Metaphysics of Ethics*, trans. E. Semple. Edinburgh: Thomas Clark.
- Kant, I. 2002. *Groundwork of the Metaphysics of Morals*, trans. and ed. T. Hill and A. Zweig. Oxford: Oxford University Press.
- Keynes, J. M. 1900. 'Obituary: Henry Sidgwick', *Economic Journal* 10: 585–91.
- Kloppenber, J. 1992. 'Rethinking Tradition: Sidgwick and the Philosophy of the Via Media', in Schultz (ed.), 369–96.
- Knobe, J. 2003. 'Intentional Action and Side Effects in Ordinary Language', *Analysis* 63: 190–4.
- Kolodny, N. 2005. 'Why be Rational?', *Mind* 114: 509–63.
- Korsgaard, C. 1986. 'Skepticism about Practical Reason', *Journal of Philosophy* 83: 5–25.
- Lacey, A. R. 1959. 'Sidgwick's Ethical Maxims', *Philosophy* 34: 217–28.
- Langenus, W. 2000. 'The Impossibility of Sidgwick's Proof', *History of Philosophy Quarterly* 17: 99–120.
- Libet, B., C. Gleason, E. Wright, and D. Pearl 1983. 'Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-potential): The Unconscious Initiation of a Freely Voluntary Act', *Brain* 106: 623–42.
- Lillehammer, H. 2010. 'Methods of Ethics and the Descent of Man: Darwin and Sidgwick on Ethics and Evolution', *Biology and Philosophy* 25: 361–78.
- Lipkin, R. 1967. 'In Defense of Sidgwick', *Philosophical Studies* 18: 70–2.
- Locke, D. 1968. 'The Trivializability of Universalizability', *Philosophical Review* 77: 25–44.
- Locke, J. 1988. *Two Treatises of Government*, ed. P. Laslett. Cambridge: Cambridge University Press.
- Lyons, D. 1965. *Forms and Limits of Utilitarianism*. Oxford: Clarendon Press.
- McCarthy, J. E. 1954. *A Consideration of the Methods of Ethics by Henry Sidgwick*. Stanford: Dept. of Philosophy.
- Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin.
- Mackie, J. L. 1992. 'Sidgwick's Pessimism', repr. in Schultz (ed.), 163–74.
- McLeod, O. 2000. 'What is Sidgwick's Dualism of the Practical Reason?', *Pacific Philosophical Quarterly* 81: 273–90.

- McMahan, J. 2002. *The Ethics of Killing: Problems at the Margins of Life*. Oxford: Oxford University Press.
- McTaggart, J. M. E. 1906. 'The Ethics of Henry Sidgwick', *Quarterly Review* 205: 398–419.
- McTaggart, J. M. E. 1927. *The Nature of Existence*. Cambridge: Cambridge University Press.
- Martineau, J. 1885. 'Professor Sidgwick on "Types of Ethical Theory"', *Mind* O.S. 10: 628–40.
- Melitz, J. 1979. 'Sidgwick's Theory of International Values', *Economic Journal* 73: 431–41.
- Mill, J. S. 1961–91. *Collected Works*, ed. J. Robson. Toronto: University of Toronto Press.
- Mill, J. S. 1998. *Utilitarianism*, ed. R. Crisp. Oxford: Oxford University Press.
- Monro, D. H. 1950. 'In Defense of Hedonism', *Ethics* 60: 285–91.
- Moore, G. E. 1903. *Principia Ethica*. Cambridge: Cambridge University Press.
- Mulgan, T. 2008. 'Sidgwick, Origen, and the Reconciliation of Egoism and Morality', *Etica & Politica* 10: 42–71.
- Mullins, T. Y. 1963. 'Sidgwick's Concept of Ethical Science', *Journal of the History of Ideas* 24: 584–8.
- Myers, F. W. H. 1904. 'Henry Sidgwick', in *Fragments of Prose and Poetry*, ed. E. Myers. London: Longmans, Green, 95–114.
- Nagel, T. 1976. 'Moral Luck', *Proceedings of the Aristotelian Society*, suppl., vol. 50: 137–51.
- Nagel, T. 1986. *The View from Nowhere*. New York: Oxford University Press.
- Nakai, D. 2011. 'Sidgwick's Utilitarianism and Practical Philosophy', in Bucolo, Crisp, and Schultz (eds), 346–92.
- Nakano-Okuno, M. 2007. 'Sidgwick and Kant: On the So-called "Discrepancies" between Utilitarian and Kantian Ethics', in Bucolo, Crisp, and Schultz (eds), 260–332.
- Nakano-Okuno, M. 2011. *Sidgwick and Contemporary Utilitarianism*. Basingstoke: Palgrave Macmillan.
- Newell, D. 1979. 'Sidgwick's Common Sense Realism', *Philosophy Research Archives* 5: 472–92.
- Nozick, R. 1974. *Anarchy, State, and Utopia*. Oxford: Blackwell.
- Nozick, R. 1989. *The Examined Life*. New York: Simon & Schuster.
- O'Neill, O. 2001. 'Sidgwick on Practical Reason', in Harrison (ed.), 83–9.
- Orsi, F. 2008. 'The Dualism of the Practical Reason', *Etica & Politica* 10: 19–41.
- Osterberg, J. 1988. *Self and Others: A Study of Ethical Egoism*. New York: Springer.
- Parfit, D. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
- Parfit, D. 2011. *On What Matters*, 2 vols. Oxford: Oxford University Press.
- Parfit, D. 2014. 'Impartial-Reason Act Consequentialism', unpub. TS.
- Parfit, D. forthcoming. 'Reply to my critics', in P. Singer (ed.), *Does Anything Really Matter? Responses to Parfit*. Oxford: Oxford University Press.

- Paton, H. J. 1971. *The Categorical Imperative: A Study in Kant's Moral Philosophy*. Philadelphia: University of Pennsylvania Press.
- Paul, S. 2009. 'How We Know What We're Doing', *Philosophers' Imprint* 9: 1–24.
- Phillips, D. 1998. 'Sidgwick, Dualism, and Indeterminacy in Practical Reasoning', *History of Philosophy Quarterly* 15: 57–78.
- Phillips, D. 2001. 'Gert, Sidgwick, and Hybrid Theories of Rationality', *Journal of Value Inquiry* 35: 439–48.
- Phillips, D. 2011. *Sidgwickian Ethics*. New York: Oxford University Press.
- Phillips, D. 2013. 'Replies to Crisp, Shaver and Skelton', *Revue d'études benthamiennes* [online] 12. <<http://etudes-benthamiennes.revues.org/676>>.
- Plato 2003. *Republica*, ed. S. Slings. Oxford: Clarendon Press.
- Prior, A. N. 1963. *Logic and the Basis of Ethics*. Oxford: Clarendon Press.
- Rabinowicz, W. and T. Rønnow-Rasmussen 2004. 'The Strike of the Demon: On Fitting Pro-attitudes and Value', *Ethics* 114: 391–423.
- Raphael, D. D. 1974. 'Sidgwick on Intuitionism', *Monist* 58: 405–19.
- Rashdall, H. 1885. 'Professor Sidgwick's Utilitarianism', *Mind* O.S. 10: 200–26.
- Rashdall, H. 1907. *The Theory of Good and Evil: A Treatise on Moral Philosophy*, 2 vols. Oxford: Clarendon Press.
- Rawls, J. 1999. *A Theory of Justice*, rev. edn. Oxford: Oxford University Press.
- Rawls, J. 2008. *Lectures on the History of Political Philosophy*. Cambridge, MA: Harvard University Press.
- Rée, J. 2001. 'Ethics, Utilitarianism, and Positive Boredom', in Harrison (ed.), 51–7.
- Reichlin, M. 2008. 'Ordinary Moral Knowledge and Philosophical Ethics in Kant and Sidgwick', *Etica & Politica* 10: 109–36.
- Richardson, H. 1991. 'Commensurability as a Prerequisite of Rational Choice: An Examination of Sidgwick's Position', *History of Philosophy Quarterly* 8: 181–97.
- Ritchie, D. G. 1892. 'Review of Henry Sidgwick, *The Elements of Politics*', *International Journal of Ethics* 2: 254–7.
- Roberts, G. 1969. 'A Fundamental Contradiction in Sidgwick's Ethics', *Critica* 3.9: 59–75.
- Ross, W. D. 1930. *The Right and the Good*. Oxford: Clarendon Press.
- Ross, W. D. 1939. *Foundations of Ethics*. Oxford: Clarendon Press.
- Ryan, A. 2007. 'The Point of View of the Universe: Mill to Dewey', in Bucolo, Crisp, and Schultz (eds), 336–64.
- Scanlon, T. M. 1998. *What Do We Owe to Each Other?* Cambridge, MA: Belknap.
- Scanlon, T. M. 2014. *Being Realistic about Reasons*. Oxford: Oxford University Press.
- Schneewind, J. 1963. 'First Principles and Common Sense Morality in Sidgwick's Ethics', *Archiv für Geschichte der Philosophie* 45: 137–56.
- Schneewind, J. 1967. 'Henry Sidgwick (1838–1900)', in P. Edwards (ed.), *Encyclopedia of Philosophy*. New York: Macmillan, 6.434–6.
- Schneewind, J. 1977. *Sidgwick's Ethics and Victorian Moral Philosophy*. Oxford: Clarendon Press.

- Schneewind, J. 1992. 'Sidgwick and the Cambridge Moralists', in Schultz (ed.), 93–122.
- Schneewind, J. 2011. 'Skorupski on Impartiality', in Bucolo, Crisp, and Schultz (eds), 394–410.
- Schofield, P. 2011. 'Sidgwick on Bentham: The "Double Aspect" of Utilitarianism', in Bucolo, Crisp, and Schultz (eds), 412–68.
- Schultz, B. 1992a. 'Henry Sidgwick Today', in Schultz (ed.), 1–61.
- Schultz, B. (ed.) 1992b. *Essays on Henry Sidgwick*. Cambridge: Cambridge University Press.
- Schultz, B. 2000. 'Sidgwick's Feminism', *Utilitas* 12: 379–401.
- Schultz, B. 2002. 'The Eye of the Universe: Sidgwick and the Problem Public', *Utilitas* 14: 155–88.
- Schultz, B. 2004. *The Eye of the Universe*. Cambridge: Cambridge University Press.
- Schultz, B. 2005. 'Sidgwick's Racism', in B. Schultz and G. Varouxakis (eds), *Utilitarianism and Empire*. Lanham, MD: Lexington, 211–50.
- Schultz, B. 2007a. 'Mill and Sidgwick, Imperialism and Racism', *Utilitas* 19: 104–30.
- Schultz, B. 2007b. 'Why Read Sidgwick Today?', in Bucolo, Crisp, and Schultz (eds), 368–406.
- Schultz, B. 2011. 'Sidgwick the Educator: Reflections on the Current Relevance of Henry Sidgwick's Philosophy of Education', in Bucolo, Crisp, and Schultz (eds), 470–560.
- Schultz, B. 2012. 'Henry Sidgwick', *Stanford Encyclopedia of Philosophy* (Fall 2012 edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2012/entries/sidgwick/>>.
- Schultz, R., 1975. 'Sidgwick on Proof in Ethics', *Philosophy Research Archives* 1: 336–51.
- Seth, J. 1896. 'Is Pleasure the Summum Bonum?', *International Journal of Ethics* 6: 409–24.
- Seth, J. 1901. 'The Ethical System of Henry Sidgwick', *Mind* N.S. 10: 172–87.
- Sextus Empiricus 1994. *Outlines of Scepticism*, trans. J. Annas and J. Barnes. Cambridge: Cambridge University Press.
- Shaida, S. 1974. 'Moore's Evaluation of Sidgwick's Hedonism', *Indian Philosophical Quarterly* 1: 112–23.
- Shaver, R. 1997. 'Sidgwick's False Friends', *Ethics* 107: 314–20.
- Shaver, R. 1999. *Rational Egoism: A Selective and Critical History*. Cambridge: Cambridge University Press.
- Shaver, R. 2000. 'Sidgwick's Minimal Metaethics', *Utilitas* 12: 261–77.
- Shaver, R. 2006. 'Sidgwick on Moral Motivation', *Philosophers' Imprint* 6: 1–14.
- Shaver, R. 2008. 'Sidgwick on Virtue', *Etica & Politica* 10: 210–29.
- Shaver, R. 2011a. 'Review of A. Hills, *The Beloved Self*', *Philosophical Quarterly* 61: 658–60.

- Shaver, R. 2011b. 'Sidgwick's Axioms and Common Sense Morality', in Bucolo, Crisp, and Schultz (eds), 562–91.
- Shaver, R. 2013. 'Utilitarianism and Egoism in *Sidgwickian Ethics*', *Revue d'études benthamiennes* [online] 12. <<http://etudes-benthamiennes.revues.org/673>>.
- Shaver, R. 2014. 'Sidgwick's Axioms and Consequentialism', *Philosophical Review* 123: 173–204.
- Sidgwick, A. and E. M. Sidgwick 1906. *Henry Sidgwick: A Memoir*. London: Macmillan.
- Sidgwick, H. 1904. *Miscellaneous Essays and Addresses*. London: Macmillan.
- Sidgwick, H. 1998. *Practical Ethics*, intro. S. Bok. New York: Oxford University Press.
- Sidgwick, H. 2000. *Essays on Ethics and Method*, ed. M. Singer. Oxford: Clarendon Press.
- Sidgwick, H. 2010–15. *The Methods of Ethics*, 6th edn, ed. J. Bennett. <<http://www.earlymoderntexts.com/pdfs/sidgwick1874.pdf>>.
- Simmons, A. J. 1982. 'Utilitarianism and Unconscious Utilitarianism', in H. Miller and W. Williams (eds), *The Limits of Utilitarianism*. Minneapolis: University of Minnesota Press, 86–100.
- Singer, M. 1974. 'The Many Methods of Sidgwick's Ethics', *Monist* 58: 420–48.
- Singer, M. 1982. 'Schneewind's Sidgwick', *Nous* 16: 339–50.
- Singer, M. 1986. 'Common Sense and Paradox in Sidgwick's Ethics', *History of Philosophy Quarterly* 3: 65–78.
- Singer, M. 1992. 'Sidgwick and Nineteenth-century British Ethical Thought', in Schultz (ed.), 65–91.
- Singer, M. 2000. 'Introduction: The Philosophy of Henry Sidgwick', in H. Sidgwick, *Essays on Ethics and Method*, ed. M. Singer. New York: Oxford University Press, xiv–xxxviii.
- Singer, P. 1974. 'Sidgwick and Reflective Equilibrium', *Monist* 58: 490–517.
- Sinha, A. P. 1979. 'On Sidgwick's Reconciliation of Ethical Theories', *Indian Philosophical Quarterly* 6: 149–58.
- Skelton, A. 2002. 'Rethinking Sidgwick: New Books on Henry Sidgwick', *International Journal of Philosophical Studies* 10: 209–17.
- Skelton, A. 2006. 'Henry Sidgwick's Practical Ethics: A Defence', *Utilitas* 18: 199–217.
- Skelton, A. 2007. 'Schultz's Sidgwick', *Utilitas* 19: 91–103.
- Skelton, A. 2008. 'Sidgwick's Philosophical Intuitions', *Etica & Politica* 10: 185–209.
- Skelton, A. 2010a. 'Henry Sidgwick's Moral Epistemology', *Journal of the History of Philosophy* 48: 491–520.
- Skelton, A. 2010b. 'On Sidgwick's Demise: A Reply to Professor Deigh', *Utilitas* 22: 70–7.
- Skelton, A. 2011. 'Utilitarian Practical Ethics: Sidgwick and Singer', in Bucolo, Crisp, and Schultz (eds), 592–632.

- Skelton, A. 2013. 'Sidgwick's Argument for Utilitarianism and his Moral Epistemology: A Reply to David Phillips', *Revue d'études benthamiennes* [online] 12. <<http://etudes-benthamiennes.revues.org/675>>.
- Skorupski, J. 1979. 'Critical Study: Sidgwick's Ethics', *Philosophical Quarterly* 115: 158–69.
- Skorupski, J. 1999. 'Irrealist Cognitivism', *Ratio* 12: 436–59.
- Skorupski, J. 2000. 'Desire and Will in Sidgwick and Green', *Utilitas* 12: 307–28.
- Skorupski, J. 2001. 'Three Methods and a Dualism', in Harrison (ed.), 61–81.
- Skorupski, J. 2007. 'Sidgwick and the Many-sidedness of Ethics', in Bucolo, Crisp, and Schultz (eds), 410–42.
- Skorupski, J. 2010. *The Domain of Reasons*. Oxford: Oxford University Press.
- Skorupski, J. 2011. 'Impartiality: Kant, Mill, Sidgwick', in Bucolo, Crisp, and Schultz (eds), 634–78.
- Skorupski, J. 2012. 'Aristotelianism and Modernity: Terence Irwin on the Development of Ethics', *European Journal of Philosophy* 20: 312–37.
- Slater, J. 1996. 'The Importance of Henry Sidgwick', intro. to H. Sidgwick, *Works*. Bristol: Thoemmes, v–liv.
- Slote, M. 1990. 'Ethics without Free Will', *Social Theory and Practice* 16: 369–83.
- Smith, A. 1976. *The Theory of Moral Sentiments*, ed. D. Raphael and A. Macfie. Oxford: Oxford University Press.
- Smith, M. 1994. *The Moral Problem*. Oxford: Blackwell.
- Smith, M. 2009. 'Desires, Values, Reasons, and the Dualism of Practical Reason', *Ratio* 22: 98–125.
- Sorley, W. 1901. 'Henry Sidgwick', *International Journal of Ethics* 11: 168–74.
- Stephen, L. 1875. 'Sidgwick's *Methods of Ethics*', *Fraser's Magazine* 11 (63): 306–25.
- Stephen, L. 1901a. 'Henry Sidgwick', *Mind* N.S. 10: 1–17.
- Stephen, L. 1901b. 'Henry Sidgwick', *Dictionary of National Biography*, suppl., vol. 3: 342–5.
- Strawson, G. 2010. 'Your Move: The Maze of Free Will', *New York Times* <<http://opinionator.blogs.nytimes.com/2010/07/22/your-move-the-maze-of-free-will/>>.
- Striker, G. 2004. 'Historical Reflections on Classical Pyrrhonism and Neo-Pyrrhonism', in W. Sinnott-Armstrong (ed.), *Pyrrhonian Scepticism*. New York: Oxford University Press, 13–24.
- Sumner, L. W. 1996. *Welfare, Happiness, and Pleasure*. Oxford: Clarendon Press.
- Sverdlik, S. 1985. 'Sidgwick's Methodology', *Journal of the History of Philosophy* 23: 537–53.
- Temkin, L. 2011. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. New York: Oxford University Press.
- Vaccari, A. 2008. 'Prudence and Morality in Butler, Sidgwick, and Parfit', *Etica & Politica* 10: 72–108.
- Velleman, J. D. 1989. *Practical Reflection*. Princeton: Princeton University Press.

- Vigna, C. 2007. 'On "the Good" which is not the Useful', in Bucolo, Crisp, and Schultz (eds), 446–56.
- Vigna, C. 2011. 'Ethics of Politics', in Bucolo, Crisp, and Schultz (eds), 680–702.
- Vincent, A. 1999. 'Bradley and Sidgwick on Philosophical Ethics', *Collingwood Studies* 6: 110–26.
- Vittone, G. 2011. 'Sidgwick: Utilitarianism and the Emancipation of Man', in Bucolo, Crisp, and Schultz (eds), 704–24.
- Walzer, M. 1983. *Spheres of Justice: A Defense of Pluralism and Equality*. New York: Basic Books.
- Wardle, J. 1983. 'Miss Anscombe on Sidgwick's View of Humility', *Philosophy* 58: 389–91.
- Watson, J. 1876. 'Hedonism and Utilitarianism', *Journal of Speculative Philosophy* 10: 271–90.
- Weinstein, D. 2000. 'Deductive Hedonism and the Anxiety of Influence', *Utilitas* 12: 329–46.
- Westbrook, D., H. Kennerley, and J. Kirk 2011. *Cognitive Behaviour Therapy: Skills and Applications*, 2nd edn. London: Sage.
- Whewell, W. 1845. *The Elements of Morality, Including Polity*, 2 vols. London: John W. Parker.
- White, N. 1992. 'The Attractive and the Imperative: Sidgwick's View of Greek Ethics', in Schultz (ed.), 311–30.
- Wiggins, D. 1995. 'Categorical Requirements: Kant and Hume on the Idea of Duty', in R. Hursthouse, G. Lawrence, and W. Quinn (eds), *Virtues and Reasons: Philippa Foot and Moral Theory*. Oxford: Clarendon Press, 297–330.
- Williams, B. 1981a. 'Moral Luck', repr. in his *Moral Luck*, Cambridge: Cambridge University Press, 20–39.
- Williams, B. 1981b. 'Persons, Character, and Morality', repr. in his *Moral Luck*, Cambridge: Cambridge University Press, 1–19.
- Williams, B. 1981c. 'Justice as a Virtue', repr. in his *Moral Luck*, Cambridge: Cambridge University Press, 83–93.
- Williams, B. 1995. 'The Point of View of the Universe: Sidgwick and the Ambitions of Ethics', repr. in his *Making Sense of Humanity*. Cambridge: Cambridge University Press, 153–71.
- Wittgenstein, L. 1967. *Philosophical Investigations*, trans. E. Anscombe, 3rd edn. Oxford: Blackwell.

Index

- anarchism 28
animals, non-human 204
Aristotle 106, 113–14
 and dialectic 83, 106–7, 141,
 209–10, 212
 doctrine of the mean of 58, 131,
 138–9, 151
 on ethics as practical 21
 influence on Sidgwick of 2, 127
 and moral motivation 218
 pluralism of ix, 159–60, 193–4
 and practical judgement 134,
 144, 146
 and priority of cognition 212
 and voluntariness 34, 147–9
axioms, self-evident 115–26

Bain, A. 56, 65–6, 72
basic actions 43
benevolence 22, 190, 214, 220
 and evolution 199
 and punishment 55
 Sidgwick's principle of 111, 119–26,
 195, 208–10
 and sympathy 73
 as a virtue 152–61
Bentham, J. 12, 71, 130, 161, 199–200, 203,
 205–6
blame 148, 218
Bradley, F.H. 17
Brink, D. 229–30
Broad, C.D. vii, x
buck-passing account of goodness 57–8,
 61–2
Butler, J. 60, 115
 and dualism of practical reason 2, 25,
 130, 234
 on pleasure and desire 72

certainty, conditions for highest 108–12
character, value of 127
chivalry 183
clarity-non-dissensus dilemma 188–90
Clarke, S. 125, 174
cognitive behaviour therapy 44
cognitivism 5–7, 97
coherentism 98, 106–7
common-sense morality 23, 145–94
compassion 159
conceptual minimalism ix, 17–19, 43–4
conformity to nature 9–10
consent theory 172–3
consequentialism 27, 51
controlling and completing arguments 211
Cosmos of Duty 233
courage 181, 186–7
Cremaschi, S. 3
Cumberland, R. 125, 215

Davidson, D. 43
Deigh, J. 14, 19
de Lazari-Radek, K. and P. Singer 3, 68,
 118, 199
demandingness 160, 226
desert 166–71
determinism 45–56, 168–9
discontinuity 86–7, 205
disinterested action 33–4
divine will 8–9
dualism of practical reason 2, 74,
 199–201, 227–34
 C.D. Broad on 227–8
 and Butler 25
 and practical judgement 4
 and Victorian common sense 184
duties, self-regarding and social 177–8,
 181, 199–200

- egoism 23–4, 195–201
equality 205–7
ethics
 as attractive or imperative 18, 57–8
 autonomy of 5–15
 as jural or teleological 20,
 102–3, 193
 methods of 21–8, 208
Euthyphro dilemma 8
expensive tastes objection 169
- flow 73
foundationalism 98, 106–7, 211–12
Frankena, W.K. viii, 6, 61, 228
free will 32–56
freedom 164–6
- Gibbard, A. 7
God 8–9, 13, 17, 21, 52–3
Golden Rule 116
good, the 57–63
 and physical processes 79
 zoological conceptions of 79
gratitude 159, 168–9
Green, T.H. 93–4
Grote, G. 103
Gurney, E. 86–7
- Hamilton, W. 92
Hare, R.M. 7
Hayward, F. 161, 196
hedonism
 deductive 92–3
 egoistic vs. global 77
 empirical 85–9
 ethical 1–2
 intuitionistic 82–3, 209–10
 objections to 93–6
 objective 90–1
 paradox of 73
 psychological 1–2, 70–5, 203–4
 synonymic 75–6
 welfare 76–84
Hobbes, T. 10, 125, 164
Hume, D.
 on objects of moral appraisal 40
 on personal identity 197
 on role of reason in ethics 11–15
 on utility of virtue 215
- humility 187–8
Hurka, T. 17, 22, 108
Hurley, P. 18
- ideals in ethics 30
impartiality 4–5, 26–7, 153, 182, 214
 ethical 123–5, 191–2, 205–6
imperatives, hypothetical and categorical
 11, 14–15, 125–6, 224–5
inductivism 98
intention 34, 39–42, 202
 and motive 127–32
internalism, motivational 19–20
intuition, faculty of 98
intuitionism
 aesthetic 101, 114, 193, 231
 dogmatic 101–3
 perceptual 99–101, 114, 193
 philosophical 103–7
 Sidgwick's conception of 1–2, 5–6, 22,
 26, 97–106
 in Whewell 1
Irwin, T. 119
- judgement, role in ethics of ix, 4, 114,
 134, 144, 145–51, 192–3, 231
judgement, suspension of ix, 84, 110–12,
 125–6
justice 29
 in Aristotle 4
 and benevolence 154
 distributive 197
 as sham-axiom 112–13
 Sidgwick's principle of 115–17, 125–6,
 157, 207
 and veracity 178
 as a virtue 161–77
- Kagan, S. 202, 203
Kant, I. 2, 12, 30, 125–6
 on categorical imperative 224–5
 on freedom 37–8
 on moral worth 130, 218
 and veracity 180–1
Knobe effect 163
Korsgaard, C. 14
- law and morality 29–30
Lecky, W. 217–18

- liberality 182–3
 libertarianism 48–56
 Lockean proviso 165–6
 luck egalitarianism 168–9
- malevolence 181–2
 Malthusianism 204
 market value 169
 marriage 158
 Martineau, J. 132
 McTaggart, J.M.E. 205
 methods 21–8, 208
Methods of Ethics, The
 reputation of vii–viii
 role of book 3 of ix–x
 Mill, J.S. vii
 on desire 71–2
 on justice 174
 on principles 3
 proof of utility of 203–4, 209, 210
 on quality of pleasures 85
 Moore, G.E. viii, 6, 11–12, 76
 and beautiful universe 24, 83
 on egoism 58, 199
 moral luck 41
 moral value 217
 motivation, moral 15–20, 150, 218
 motive 129–32
 Mullins, T. viii
 Myers, F. viii, 234
- Nagel, T. 198
 Nakano-Okuno, M. 118
 non-cognitivism 7
 non-identity problem 166
 non-naturalism 5–7, 11–12
 Nozick, R. 95, 164, 165
- Oakeshott, M. 161
 ought 7–17
- parapsychology 234
 Parfit, D. vii, viii, 196–7, 205, 210, 230
 perfectionism 10, 24–6, 51–2, 82–3,
 95–6, 153
 personal identity 197
 personal point of view 197–8
 Phillips, D. 6, 7, 11–12, 15, 108, 189,
 212, 230
- Plato 25, 31, 113, 200
 on pleasure 88, 89
 pleasure
 apprehension view of 68–9
 definition of 63–70
 desirable consciousness view of
 69–70
 feeling-tone view of 66–8
 internalist and externalist accounts
 of 66–7
 nineteenth-century views on
 92–3
 volitional view of 63–6
 political obligation 171–4
 politics 28–31
 population 204–5
 positive psychology 87
 precision in ethics 3–6, 121–2
 Price, R. 6
 pro tanto ('prima facie') principles 22,
 23, 192–4, 214, 231
 problem of evil 233
 promises 174–6, 178–9, 215–16
 property 165–8
 prudence 178
 and morality 36
 Sidgwick's principle of 117–19, 122–3,
 125–6, 195–6
 as a virtue 184–6
 publicity 225–6
 punishment 54–6, 167, 170–1
 purity 185–6
- quantum theory 46
 quietism, metaphysical 5–7, 98
- Rashdall, H. 196
 rationality
 and freedom 34–6
 ought and 11–15, 19–20
 Rawls, J. viii, 61, 106, 193–4
 realism, moral 6–7
 repugnant conclusion 205
 revelation 232–3
 reverence 157–8
 rightness, subjective and objective 78,
 131, 132–4, 151, 201–3, 219
 rights, natural 164–6
 Ritchie, D.G. 223

- Ross, W.D. viii, 106, 192–4, 215–16
 pluralism of ix, 22, 231
 on value of virtue 143–4
- Rousseau, J.J. 173
 rule utilitarianism 171, 223
- sanctions 13, 199–200
- Sartre, J.-P. 44
- Schneewind, J. 5, 18, 21, 26, 60,
 106, 117, 118, 120, 121–2, 208,
 210, 213
- self-control 145–51, 184–6
- self-effacement 185, 191
- self-evidence 107–12, 155–6, 176, 188–9
- self-realization 10
- self-regarding virtue 184–6
- separateness of persons 124, 196–7
- Shaftesbury, Lord 215
- sham-axioms 112–15, 150
- Shaver, R. 7, 16, 18, 19, 119, 122–3,
 196–7
- Sidgwick, H.
 intellectual development of 1–3
 on race 155
 resignation of fellowship of 179
 style of x
 view of theism of 233–4
- Singer, P. 3, 106
- Skelton, A. viii, 105
- Skorupski, J. 228
- Smith, A. 215
- socialism 170
- Socrates vii, ix, 113
- special relations 155–61
- Spencer, H. 30, 65, 92–3, 164, 166, 223
- Spinoza, B. 10
- Stephen, L. 12, 223
- Stoics 114–15
- Stout, G.F. 92
- Sumner, L.W. 66
 supererogation 136–8, 154–5, 183, 226–7
- temperance 145–51, 184–6
- Tennyson, Alfred Lord 63
- unconscious utilitarianism 215–16,
 219–20
- under-determination objection 190
- universalizability 2, 131
- universe, point of view of 122–3, 198
- utilitarianism 23, 201–27
 and intention 127–9
 and justice 176–7
- validity 6, 210–13
- veracity 174, 177–81
- virtue 127–94
 cultivation of 142–4
 and duty 135–6
 and emotion 138–40
 and good of the agent 20, 25–6,
 78–80, 83
 and motivation 140–2
- virtue ethics 127, 132, 137, 141, 143
- voluntariness 38–42, 147–9
- Weinstein, D. 92
- Whewell, W. 1, 3, 193
- will, the 38–45
- Williams, B. 12, 139, 140
 on agent-regret 40
 on one thought too many 20, 140
 on personal point of view 197–8
 on the Socratic question ix
- wisdom 145–51, 184, 191
- Wollaston, W. 174
- women, extension of franchise to 173
- wrong kind of reasons objection 61