



OXFORD

TAKING UTILITARIANISM SERIOUSLY

Christopher Woodard

Taking Utilitarianism Seriously

Taking Utilitarianism Seriously

CHRISTOPHER WOODARD

OXFORD
UNIVERSITY PRESS

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

© Christopher Woodard 2019

The moral rights of the author have been asserted

First Edition published in 2019

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: 2019941392

ISBN 978-0-19-873262-4

Printed and bound in Great Britain by
Clays Ltd, Elcograf S.p.A.

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 B, M, P, and E

Preface

Utilitarianism is often dismissed very quickly by those engaged in ethics and political philosophy. It is an unfashionable view, and it has come to be associated with philosophical and psychological crudeness. In some quarters this is traced primarily to the fact that it is a consequentialist theory. But utilitarianism has a bad name even amongst consequentialists. It is thought to be the form of consequentialism that you endorse if you are simple-minded and lack an appreciation of the psychological and ethical complexities of life. Many contemporary philosophers seem to share something of Nietzsche's view:

Just look at the indefatigable, unavoidable English utilitarians, for example, how awkwardly and honorably they walk in Bentham's footsteps... No new thoughts, no sign of any subtle change or fold in an old thought, not even a real history of the earlier thought: an *impossible* literature on the whole, unless you know how to sour it with some malice.¹

At any rate, utilitarianism often fails to get a fair hearing. The aim of this book is to improve the seriousness with which utilitarianism is considered, including by those who reject it.

The book begins by identifying six common philosophical objections to utilitarianism. It then seeks to develop a version of utilitarianism that is equipped to answer them. In comparison to some other versions of utilitarianism, the one developed here is both broader in the topics it addresses and more complex in its structure. I try to show that utilitarianism has the resources to match and explain the complexity of our ethical ideas.

One result is that the book discusses many topics in moral and political philosophy. Utilitarianism is often presented, simply, as a theory of the rightness of actions. This book considers it, in addition, as a theory of reasons for action, of moral rights, of justice and the importance of equality, of legitimacy and democracy, and of virtuous agency. The underlying idea is

¹ Nietzsche 2002/1886: 119, emphasis in the original.

that we get a better grip on the prospects of an ethical theory if it is developed in some detail.

My hope is that the book will be interesting and useful for specialists, graduate students, and advanced undergraduates studying utilitarianism in courses on ethics and political philosophy. It is not intended to be a balanced discussion of existing views, but instead to develop a new and attractive version of utilitarianism. Nevertheless, I have tried to trace connections with existing views in the footnotes, and to explain some different ways in which utilitarian ideas may be developed where appropriate. In this way, I hope that the book may help others to appreciate the untapped resources of the utilitarian tradition, and prompt them to make their own contributions to its further development.

I could not have written this book without research leave. I am very grateful to the University of Nottingham for a period of leave in 2013–14 during which I began work on the book, and for another period of leave in 2016–17 which enabled me to write the first half of it. I am also very grateful to the Leverhulme Trust for generously funding a Research Fellowship in 2017, during which I finished a complete draft.

I benefited greatly from the opportunity to discuss this material with three very talented groups of students. I was fortunate to be able to develop my ideas in the classroom, and I learned a great deal from my students' excellent questions and comments. I am very grateful to all of them. I am also very grateful to colleagues at Nottingham in Philosophy and in Politics for encouragement and discussion.

Two academic societies have provided me with intellectual kinship and continuing education. I have learned an enormous amount from papers delivered at meetings of the *British Society for Ethical Theory* and of the *International Society for Utilitarian Studies*, and from discussions with many colleagues and friends at these meetings.

For helpful comments or conversations I would like to thank Kaoru Ando, Tom Baker, Krister Bykvist, Bruce Chapman, Sophie Grace Chappell, Richard Yetter Chappell, Roger Crisp, Alexander Dietz, Nir Eyal, Guy Fletcher, Alex Gregory, Katharine Jenkins, Ian Kidd, Joseph Kisolo-Ssonko, Rob Lawlor, Uri Leibowitz, Penelope Mackie, Greg Mason, Brian McElwee, Peter Niesen, Douglas Portmore, Carolyn Price, Janosch Prinz, Peter Railton, Nicholas Shackel, Hillel Steiner, Bart Streumer, Aness Webster, Sam Wren-Lewis, and Michael Zimmerman. As will become clear in Chapter 5, Alexander Dietz's comments proved to be very influential on

the claims made in this book. I would also like to thank audiences in Bloomington, Boulder, Karlsruhe, Leeds, Lille, Lucca, Manchester, Milton Keynes, Norwich, Nottingham, Oxford, Rennes, Swansea, Vienna, Warwick, and Yokohama for their comments and questions.

I could not hope for a better mixture of encouragement and astute criticism than I received from my colleagues Zachary Hoskins and Neil Sinclair in detailed comments on a substantial portion of the first draft. I am very grateful to both of them. For encouragement and inspiration over many years I am extremely grateful to Brad Hooker, Tim Mulgan, and the late Susan Hurley. I have learned a great deal from each of them and their writings, and each has been extremely generous with their help.

I am most grateful to Peter Momtchiloff for having confidence in this project from the very beginning, and for his deft hand in guiding it along ever since. I am grateful also to his colleagues at Oxford University Press for their work on the book, and especially to Susan Frampton for her careful copy-editing. Two readers for Oxford University Press provided very detailed and helpful comments which enabled me to avoid several errors and to improve the final version significantly. I am very grateful to them for their care and generosity in providing these comments.

Chapter 3 of this book features some of Shelly Kagan's diagrams of forms of consequentialism. These originally appeared in Shelly Kagan, 'Evaluative Focal Points', in B. Hooker, E. Mason, and D. Miller (eds) *Morality, Rules, and Consequences: A Critical Reader*. Edinburgh: Edinburgh University Press, Copyright © 2000: 134–55. I am very grateful to Edinburgh University Press for granting permission to use them.

Finally, I am most grateful to my family for their love and patience during the long period in which this book consumed most of my energy. It is dedicated to them.

Contents

| | |
|---|------|
| <i>List of Table and Figures</i> | xiii |
| 1. Introduction | 1 |
| 1.1 What Is Utilitarianism? | 3 |
| 1.2 What Is to Come | 7 |
| 2. Six Objections | 13 |
| 2.1 Pig Philosophy | 14 |
| 2.2 Abhorrent Actions | 18 |
| 2.3 Demandingness | 21 |
| 2.4 Separateness of Persons | 25 |
| 2.5 Politics | 29 |
| 2.6 Psychology | 33 |
| 2.7 Conclusion | 36 |
| 3. Basic Ideas | 37 |
| 3.1 Reasons | 37 |
| 3.2 Rightness | 43 |
| 3.3 Two Ways to Avoid Fragmentation | 47 |
| 3.4 Three Ways to Accommodate Fragmentation | 52 |
| 3.5 Utilitarian Theories of Reasons | 57 |
| 3.6 Conclusion | 62 |
| 4. Well-Being | 63 |
| 4.1 Philosophical Theories of Well-Being | 64 |
| 4.2 What We Know about Well-Being | 71 |
| 4.3 Alienation as Evidence | 73 |
| 4.4 Changing Values | 77 |
| 4.5 Discovering What Is Good for You | 80 |
| 4.6 Promoting Well-Being | 82 |
| 4.7 Conclusion | 84 |
| 5. Two Kinds of Reasons | 85 |
| 5.1 Act Consequentialism | 86 |
| 5.2 Pluralism | 89 |
| 5.3 The Minimal Constraint on Eligibility | 97 |
| 5.4 Rule Consequentialism | 100 |

| | | |
|-----|--|-----|
| 5.5 | Accepting the Willingness Requirement | 105 |
| 5.6 | Narrowing Eligibility | 108 |
| 5.7 | Conclusion | 112 |
| 6. | Moral Rights | 113 |
| 6.1 | The Concept of Moral Rights | 114 |
| 6.2 | Existing Utilitarian Theories of Moral Rights | 118 |
| 6.3 | A Broader Indirect Theory | 123 |
| 6.4 | The Benefits of Respect for Moral Rights | 126 |
| 6.5 | The Contingency of Moral Rights | 132 |
| 6.6 | Conclusion | 135 |
| 7. | Justice and Equality | 137 |
| 7.1 | Distributive Justice | 138 |
| 7.2 | Justice for Utilitarians | 141 |
| 7.3 | Kinds of Equality | 146 |
| 7.4 | Utilitarianism and Substantive Equality | 151 |
| 7.5 | Known Expensive Needs | 157 |
| 7.6 | Conclusion | 161 |
| 8. | Legitimacy and Democracy | 163 |
| 8.1 | Government House Utilitarianism | 164 |
| 8.2 | Democracy as Eliciting and Aggregating Preferences | 168 |
| 8.3 | Legitimacy and its Political Importance | 170 |
| 8.4 | Utilitarianism and Legitimacy | 174 |
| 8.5 | Should Utilitarians Be Democrats? | 180 |
| 8.6 | Conclusion | 185 |
| 9. | Virtuous Agents | 186 |
| 9.1 | Reasons and Rightness | 188 |
| 9.2 | Cluelessness | 192 |
| 9.3 | Good Decision Procedures | 195 |
| 9.4 | Praiseworthiness | 201 |
| 9.5 | Virtue | 205 |
| 9.6 | Conclusion | 209 |
| 10. | Conclusion | 211 |
| | <i>Bibliography</i> | 219 |
| | <i>Index</i> | 239 |

List of Table and Figures

Table

| | |
|--|----|
| 3.1 Three kinds of deontic status of an action | 44 |
|--|----|

Figures

| | |
|---|-----|
| 3.1 Hooker's proposal | 48 |
| 3.2 Moore's proposal | 51 |
| 3.3 Perspectival/non-perspectival pluralism | 53 |
| 3.4 Dissociating reasons from rightness | 54 |
| 3.5 The simple picture | 55 |
| 3.6 Kagan-diagrams of Act, Rule, and Motive Utilitarianism | 58 |
| 3.7 Kagan-diagrams of Act Utilitarianism's indirect evaluation of rules and motives | 59 |
| 4.1 Sally's first choice | 78 |
| 4.2 Sally's second choice | 80 |
| 7.1 Why the simple argument does not work | 155 |

1

Introduction

In many parts of public culture utilitarianism appears to be in the ascendancy. Governments have used cost–benefit analysis to make and evaluate public policies for some time. This is a broadly utilitarian method which uses people’s preferences over outcomes to evaluate matters of great importance—such as how to allocate health or educational resources.¹ In the last ten years or so many governments have gone further, explicitly adopting the aim of measuring and promoting national well-being. In many other parts of life, too, it has become very common to frame issues in a broadly utilitarian way, as though all that matters ultimately is well-being.

However, these broadly pro-utilitarian attitudes are in stark contrast with the attitudes of many philosophers towards utilitarianism. Though there is still some support for utilitarianism within academic philosophy, many philosophers take the view that utilitarianism has been shown to be wrong by a set of well-known objections. This attitude is not universal, but it is widespread.

Yet the attitudes to utilitarianism among philosophers have another intriguing feature. One might think that, if it were so easy to see that utilitarianism is wrong, avoiding believing it would be quite straightforward. But many philosophical critics of utilitarianism emphasize its attractions. Philippa Foot wrote that ‘[i]t is remarkable how utilitarianism tends to haunt even those of us who will not believe in it. It is as if we forever feel that it must be right, although we insist that it is wrong.’² Richard Kraut wrote that ‘[i]t is . . . possible to slip into utilitarianism by sheer inattention.’³ Like a hole in the ground or a bad habit, it is worth taking the trouble to avoid slipping into utilitarian ways of thinking. The idea that utilitarianism exerts some attractive force which takes effort to resist perhaps explains the many attempts by philosophers to diagnose both the source of its appeal and the deep ways in which, supposedly, it goes wrong.⁴

¹ See Wolff 2006 and Orr 2007.

² Foot 1985: 196. Foot then attempted an ‘exorcism’.

³ Kraut 2007: 17.

⁴ See Rawls 1999a: 19–30, Williams 1973, Scheffler 1994, Foot 1985, and Thomson 1997.

This book tries to show that utilitarianism can be defended from six common philosophical objections. I attempt to do this by elaborating a version of utilitarianism in some detail, covering a broader range of topics than is usual. For example, the book discusses utilitarian views of reasons for action, rightness, moral rights, justice, equality, legitimacy, democracy, decision-making, and virtue. The purpose of discussing all of these topics is to show that utilitarianism has sufficient internal resources to be a well-rounded ethical theory, with a plausible claim to match up to the complexity of our ethical and political ideas. Though it is often presented as a very simple theory, it need not be.

The six objections are identified and discussed in Chapter 2. Subsequent chapters discuss components of the overall theory—such as, for example, what it says about reasons, moral rights, equality, and virtue. We return to the objections in Chapter 10, once the components of the theory have been explained. I then claim that this version of utilitarianism offers plausible responses to all six objections.

This exercise has two purposes. One is to motivate interest in the specific version of utilitarianism outlined in Chapters 3–9. This version has some novel features, which I hope contribute to its plausibility and interest. It is, very roughly, an indirect form of utilitarianism with some similarities to Rule Utilitarianism, and it is developed in a way that addresses topics in both moral and political philosophy. The theory is by no means complete, and I try to highlight the gaps. Still, I hope that it is developed in enough detail for us to be able to discern its character and to reach some conclusions about its plausibility.

The second purpose is more general and more important. I hope to contribute to a renewal of philosophical interest in utilitarianism. If the specific version outlined in this book is not obviously defeated by the six common objections to be discussed, then those objections do not suffice to show that utilitarianism is a dead end. If this book shows that it is possible to answer these objections, it may encourage others to develop new, better, versions of utilitarianism.

Answering these objections is in one way a modest aim. I do not aim to show that, all things considered, we have more reason to accept a version of utilitarianism than any other ethical view. That would require comparison of versions of utilitarianism with other views, and this book is mostly silent on such comparisons. Nor do I spend much time trying to make a positive case for utilitarianism, beyond noting the intuitive plausibility of its constituent ideas. Instead the main focus is on articulating a version of utilitarianism

that can be defended from the six objections identified. The book is thus largely non-comparative and largely defensive. In both of these ways it falls short of offering a full argument in support of the theory it presents.

In another way, however, the aim is not at all modest. As noted above, the theory it develops addresses many difficult topics in moral and political philosophy. To address any one of these topics involves taking a stand on many difficult questions. To address all of these topics at once may be considered foolhardy.

There are two reasons for trying to develop a version of utilitarianism that is comprehensive in this way. One is that the common objections to utilitarianism themselves cover a wide range of topics, as we shall see in Chapter 2. To address all of the objections at once requires developing a theory that addresses many topics. The second reason has to do with a broader methodological point. Reaching a verdict about a philosophical theory involves developing it and assessing its merits. How should we divide time and effort between development and assessment? Delaying assessment can lead to wasted effort exploring dead ends. On the other hand, if we assess only simple versions of theories we may overlook some of their potential. This seems to me to happen with utilitarianism. It is often discussed in its simplest and narrowest form, as if it were a single proposition about the rightness of actions. I hope to show that it is more promising when developed more fully.

1.1 What Is Utilitarianism?

As its sympathizers have often pointed out, utilitarianism suffers from some of the connotations of its name.⁵ In ordinary usage, ‘utilitarian’ suggests something that is plain or merely functional, as contrasted with something that is fine or beautiful. Some of these connotations also carry over to philosophical usage. In some contexts, describing a view as ‘utilitarian’ may suggest that it is crude or simple-minded, or at least that it does not take account of all relevant considerations.⁶ Utilitarianism is also still

⁵ J. S. Mill wrote: ‘A passing remark is all that needs be given to the ignorant blunder of supposing that those who stand up for utility as the test of right and wrong, use the term in that restricted and merely colloquial sense in which utility is opposed to pleasure. An apology is due to the philosophical opponents of utilitarianism, for even the momentary appearance of confounding them with anyone capable of so absurd a misconception...’ Mill 1998/1861: 54.

⁶ ‘The technical use of the term Utilitarianism... has never become entirely divested of certain associations connected rather with the ordinary meaning of the word “utility,” and

associated with the views of Jeremy Bentham in particular, including his views about human psychology. This does not always redound to its credit.

Thus it is important to define what sort of view we have in mind. As I shall understand it, utilitarianism is a family of ethical views united by acceptance of three doctrines.⁷ The first of these doctrines is *consequentialism*. This is the claim that we can explain ethical phenomena such as the rightness of actions, the justice of institutions, and the virtues and vices of agents in terms of the goodness of outcomes. Consequentialists think that we can explain rightness, justice, and virtue in terms of the goodness of outcomes because they think that rightness, justice, and virtue each consist in one or more relationships—specified differently by different consequentialist theories—to good outcomes.⁸

The simplest and most familiar form of consequentialism is Act Consequentialism. As usually formulated this is a view about the rightness of actions, according to which an action is right if and only if, and because, its outcome would be at least as good as that of any relevant alternative. One way in which other forms of consequentialism differ is by offering different answers to the same question. For example, Rule Consequentialists claim that what makes an action right is that it is permitted by the best set of rules. Like Act Consequentialists, they seek to explain the rightness of actions in terms of the goodness of outcomes. Unlike Act Consequentialists, however, they claim that what explains the rightness of an action is an indirect relationship between it and outcomes: the action is related to a set of rules, which are themselves said to have good outcomes. Similarly, Motive Consequentialists claim that what makes an action right is that it is the action that would be performed, in the circumstances, by someone with the best motives. Here, motives are evaluated in terms of their outcomes, and the rightness of actions is said to depend on their relationship to good motives.

Consequentialist theories may also differ from each other in the questions they address. Act Consequentialism, Rule Consequentialism, and Motive Consequentialism are all standardly presented as rival answers to the *same* question, of what makes actions right. But this is not the only question that

with the supposed practical applications of the Utilitarian theory, than with the essential logic of the theory itself' Albee 1902: xv.

⁷ My characterization of utilitarianism broadly follows that given in Sen 1979.

⁸ Consequentialists use 'outcome' in a somewhat technical way. On this usage, something's 'outcome' includes everything that would happen were that thing realized. See Chapter 2 Section 2.1, this volume.

consequentialists can try to answer. They may also seek to explain what reasons agents have, for example, or what makes an institution just, or what makes a decision good or bad. In its most ambitious forms, consequentialism claims that all of these questions can be answered in terms of the goodness of outcomes.⁹

Consequentialist theories also differ from each other in their claims about what makes one outcome better than another.¹⁰ Utilitarian theories are distinguished from other consequentialist theories by their commitment to two doctrines about the value of outcomes. The first of these is *welfarism*. According to welfarism, all well-being has noninstrumental value, and nothing else has noninstrumental value.¹¹ This means that all well-being is good independently of its effects, and nothing else is. Though other things can be good or bad, their goodness or badness depends on their contribution to well-being. Welfarists thus claim that autonomy is good insofar as it promotes well-being, and that ill-health is bad insofar as it inhibits well-being. Since nothing else has noninstrumental value, the value of an outcome depends, according to the welfarist, entirely on the well-being it contains.

Welfarism is an important part of the utilitarian's account of the value of outcomes, but it is not a complete theory of value. One respect in which it is incomplete is that it does not tell us how, exactly, the value of an outcome depends on the well-being it contains. The third defining commitment of utilitarianism, *sum-ranking*, is an answer to this question. According to this claim, the value of an outcome is the sum of the goods (and bads) existing in

⁹ 'Global Consequentialism' claims that everything should be evaluated in terms of the goodness of its outcomes. See Parfit 1987: 25, Pettit and Smith 2000, and Driver 2014. As I shall explain in Chapter 3, however, we should distinguish between two issues: what gets evaluated in a consequentialist way, and how it gets evaluated. 'Global Consequentialism' advocates consequentialist evaluation of everything, but it also advocates direct evaluation of everything, i.e. evaluating each thing in terms of its own effects. It is important to note that other forms of global consequentialist evaluation are possible.

¹⁰ Sometimes consequentialism is defined such that it is committed to an 'agent-neutral' or impartial account of the goodness of outcomes. On this definition, consequentialism is just one kind of 'teleology', distinguished from other kinds by its impartial or agent-neutral account of goodness. For example, see Scheffler 1988: 1 and 1994: 1. On other occasions consequentialism is defined more broadly, so that a view can be consequentialist even if its account of the goodness of outcomes is 'agent-relative' or partial. For example, see Portmore 2011: 34–8. On this second usage, 'consequentialism' and 'teleology' are synonyms. In this book I understand consequentialism in the broader way, drawing no distinction between 'consequentialism' and 'teleology'. Utilitarianism is, in any case, committed to an agent-neutral or impartial theory of goodness.

¹¹ Sen 1979 is a prominent criticism of welfarism. Moore and Crisp 1996 is a defence of it, including against Sen's criticisms. Keller 2009 is a recent survey of work on welfarism.

that outcome. As I am defining ‘utilitarianism’, utilitarians are committed both to welfarism and to sum-ranking. This means that they claim that one outcome is better than another just in case it contains more well-being.¹²

For the purposes of this book, ‘utilitarianism’ refers to a family of views all of which seek to explain the rightness of actions, reasons for action, justice, virtues and vices, or other ethical phenomena in terms of some relationship between the phenomenon at issue and the goodness of outcomes. Since I have defined utilitarianism such that it includes commitment to welfarism and sum-ranking, all forms of utilitarianism also claim that one outcome is better than another just in case it contains more well-being. However, utilitarian theories differ from each other in important ways. For one thing, they differ in the questions they address. Second, they differ over the explanatory relationship between, say, an action’s rightness and good outcomes. Third, they differ over what exactly well-being is, and so what exactly makes one outcome better than another. One major theme of this book is that these dimensions of variation are sufficient to allow plausible responses to the most common objections to utilitarianism. Even when defined in the relatively narrow way that I have defined it, utilitarianism is a broad tradition of ethical thought.¹³

Not all of the theories in this tradition have a feature that has sometimes seemed to be among utilitarianism’s defining traits and chief attractions. I have in mind the feature, shared by some but not all utilitarian theories, that the role of judgement between conflicting considerations is minimized, with the result that ethical issues are pictured as, in principle, amenable to being solved by calculation. This is a feature of some versions of utilitarianism, and to some people it is attractive. It seems to result if we adopt a fairly straightforward view of the nature of well-being, an optimistic view about the precision with which we can measure well-being and make interpersonal comparisons of it, and a fairly straightforward view of the structure of utilitarianism. Under these further assumptions there may be little room,

¹² Some allow other utilitarian functions from well-being to goodness, such as the claim that the value of an outcome is given by the average level of well-being it contains. The difference between these views is salient mainly with respect to population ethics: see Parfit 1987: pt. 4. This book is not primarily concerned with population ethics, and so one option would be to remain agnostic about the function from well-being to goodness. However, I restrict ‘utilitarianism’ to those theories that accept sum-ranking in order to be as clear as possible about the theory of value under discussion, and because I believe that sum-ranking is correct.

¹³ It is hard to discuss traditions of thought. James Griffin writes: ‘One cannot say anything definitive about a rich philosophical tradition in a few pages—or in a book, for that matter. Rich traditions, by definition, have too many resources for that’ Griffin 1996: 103–4.

in principle, for difficult judgements about the relative importance of competing considerations.

However, these further assumptions are not essential to utilitarianism as I have defined it, and they are not adopted by the version of utilitarianism that I will articulate in this book. According to that version there can be competing considerations, and sometimes we have to make difficult judgements about their interaction and relative importance. The result is a picture according to which ethics is much messier than a series of calculations.

1.2 What Is to Come

My arguments depend in many places on appeals to intuition. Sometimes these are intuitions about specific cases, in which relevant facts about some agent's situation and options are stipulated. On other occasions they are intuitions about abstract principles, such as consequentialism, welfarism, or sum-ranking. When I claim that some feature enables a 'plausible reply' to one of the six objections we will be considering, this boils down to the idea that it fits with judgements about cases or principles that, I take it, are intuitively plausible.

As this may suggest, I will be using the method of philosophical argument known as the search for 'reflective equilibrium'. We start with intuitive judgements about principles and cases. We then sift through these intuitions, discarding those about which we are less sure—perhaps because we suspect that they reflect biases or vested interests. Those which remain we label 'considered convictions'. We then seek a theory that would tie these convictions together, showing how the judgements about cases can be explained by plausible principles that also cohere with our wider background beliefs. Relevant background beliefs include other philosophical beliefs, such as beliefs about personal identity, plus any other belief that is logically related to the claims made by our philosophical theory.¹⁴

It would be an enormous achievement to find an ethical theory that satisfies the conditions of reflective equilibrium. At the very least, it would

¹⁴ This describes what is sometimes known as 'wide reflective equilibrium'. See Rawls 1999a: 17–18 and 42–5, Daniels 1979, and Hooker 2000: ch. 1. Hooker 2000: 16 doubts that our beliefs about non-moral matters will provide much additional constraint on our choice between moral theories. However, it seems to me that they may do so. For example, consider the discussion of possible logical connections between utilitarianism and religious beliefs in Nelson 1991, Chappell 1993, and Mulgan 2015: pt. 3.

show how we can fit our ethical convictions into a broader coherent outlook on the world. If we were to achieve it, there would of course be further questions about whether this sort of coherence suffices for warranted belief. However, we do not have to settle these further questions in order to treat reflective equilibrium as a worthwhile goal in normative ethics. Even if we were to conclude that it does not suffice for epistemic warrant, knowing which theory or theories pass the test of reflective equilibrium could not fail to be relevant to philosophical ethics. In any case, I will employ the method of reflective equilibrium in considering the plausibility of replies to the criticisms of utilitarianism.¹⁵

Of course, appeals to reflective equilibrium are always open to question. For one thing, people have different intuitions about cases.¹⁶ They also differ in their judgements about the plausibility of abstract claims, and over how to trade off the various components of coherence. Thus they are liable to reach different conclusions about the extent to which any given theory satisfies the conditions of reflective equilibrium. For this reason, any appeal to the method of reflective equilibrium involves matters of judgement.

The reasons that many philosophers do not think utilitarianism a good bet probably do not have much to do with the plausibility of its central assumptions. Each of the three defining features of utilitarianism is attractive when considered as an abstract claim. When we consider each claim in itself, and not with its implications about cases in mind, each looks like a reasonable starting assumption from which to build an ethical theory. Consequentialism is the assumption that goodness of outcomes is basic, in the sense that we can explain other ethical phenomena in terms of some relationship to good outcomes. This is, I take it, a plausible assumption; it is at least as plausible, prior to any worries about implications about cases, as its major rivals.¹⁷ Applied to reasons for action, for example, it claims that all reasons

¹⁵ Perhaps there are other appropriate methods. It might be possible to construct an ethical theory from axioms of rational choice, say, or by analysing the use of ethical language.

¹⁶ Tim Mulgan has distinguished between 'austere' and 'complacent' views of morality. On the complacent view, our common beliefs about morality are taken to be in generally good order, while on the austere view they are suspected of being 'self-serving delusions' 2015: 20–1. These differences may reflect different views about which common intuitions deserve to be treated as 'considered convictions'. However, they may also reflect different predictions about how many of our considered convictions will be preserved in reflective equilibrium. My view is that we do not yet know how austere the correct moral theory will be, since we do not yet know what the correct moral theory is. Robert Adams takes a different view, claiming that '[t]here is a limit to how far pretheoretical opinion can be revised without changing the subject entirely' 1999: 246.

¹⁷ See Ewing 1948: 103 and Scheffler 1988: 1.

for action can be explained by some relationship between actions and good outcomes.

Welfarism is also a plausible claim. It is very hard to deny that well-being—the value of a life for the subject whose life it is—is *usually* non-instrumentally good. It is hard to deny this, at least, with respect to lives that already exist. It is somewhat plausible to go further, and claim that well-being is *always* noninstrumentally good. In cases where we might doubt this, as with the pleasure a sadist gains from torturing his victim, welfarists have somewhat plausible things to say. They can say, for example, that this pleasure does not really add to his well-being, on the best understanding of well-being; or they can claim that, though it is bad that the torture occurs, given that it does occur it is better that the sadist gains from it than that he does not. Moreover, it is also somewhat plausible to claim that *only* well-being is noninstrumentally good. Of course, welfarists allow that other things are good as well; they just claim that nothing apart from well-being is noninstrumentally good. But this is, again, a somewhat plausible claim. It is not obviously false to say that the value of autonomy, say, or of beauty or knowledge, is entirely accounted for by its contribution to well-being.¹⁸

Sum-ranking is also plausible. This is the claim that the value of an outcome is the sum of the goods (and bads) existing in that outcome. It thus claims that the value of an outcome is related in a simple way to the value of the things it contains: the value of the outcome is the sum of the values of the things it contains. As with consequentialism and welfarism, this claim is plausible in the sense that it is a reasonable assumption to make when constructing a theory. It could turn out that any or all of these claims is false. But they are promising places to start. Indeed, their face value plausibility probably explains the persistent attractiveness of utilitarianism—which, as we have noted, is much remarked upon by its critics.

To the extent that utilitarianism seems like a bad bet, then, it is not because its defining assumptions are implausible on their face. Instead it is because the theories that have been constructed from these assumptions are thought to be open to one or more powerful objections. Chapter 2 identifies six such objections. These are that utilitarianism has an inadequate theory of value, that it countenances abhorrent actions, that it is too demanding of agents, that it fails to respect the separateness of persons, that it cannot explain the distinctiveness of political issues from moral issues, and that it is

¹⁸ Shaver 2004 claims that welfarism is the most attractive feature of utilitarianism.

committed to an inadequate psychology. While utilitarianism has been the target of many other objections, these six objections have been very influential. Showing how they can all be answered by a single theory will be the task for the remainder of the book.

Chapter 3 sets out the assumptions I will make about the relationships between some of the basic topics of ethical thought: normative reasons for action, rightness and wrongness, praiseworthiness and blameworthiness, and good deliberation. One important assumption is that rightness is a function of normative reasons. On this conception, the reasons for or against an action determine whether it is right or wrong. This picture also implies that a relevant explanation of why some action is right or wrong is given by citing the reasons for and against it. But acceptance of this picture raises a question of whether it is possible to maintain all of the close connections that are sometimes posited between reasons and rightness, on one hand, and praiseworthiness or blameworthiness, and good or bad deliberation, on the other.

This question is pressing because reasons and rightness, I claim, are both ‘non-perspectival’: reasons and rightness depend on the actual consequences of actions, rather than on the consequences an agent expects or which it would be reasonable for her or for someone else to expect. On the other hand, whether someone deliberated well, or is praiseworthy or blameworthy, certainly depends on how things seemed to her—which is to say that these are certainly perspectival matters. Thus the view I argue for in Chapter 3 incurs a significant cost: of explaining what connection, if any, there is between (non-perspectival) reasons and rightness, and (perspectival) good deliberation and praiseworthiness. I try to meet this challenge later in the book, in Chapter 9.

Chapter 4 discusses well-being. It begins by considering traditional philosophical debates about the nature of well-being, concluding that the central debate—between subjectivists and non-subjectivists—is currently in stalemate. This is a disappointing conclusion. However, I claim that we know enough about well-being to make reasonable judgements about the plausibility of welfarism. In addition, the chapter argues that we can use a famous idea of Peter Railton’s for epistemological purposes. Railton claims that a person’s good would not be something she finds alienating.¹⁹ Some try to defend subjectivism using this idea. As I explain in Chapter 4, I do not think

¹⁹ Railton 2003b: 47.

that these attempts work. But rather than treat Railton's idea as an insight into the metaphysical nature of well-being, we can treat alienation as evidence about which things are sources of well-being. The chapter claims, further, that one way to enable people to promote well-being is to make it easier for them to discover which things are or would be good for them. This enables us to make an epistemic utilitarian argument for the importance of liberty and security. We thus reach a conclusion about how to promote well-being via an epistemic route, despite failing to resolve the central question about the nature of well-being.

Chapter 5 discusses two kinds of reasons. It argues for a pluralist theory of reasons, according to which there are 'act-based' reasons (reasons to make outcomes better) and 'pattern-based' reasons (reasons to contribute to good patterns of action). It explains how the concept of pattern-based reasons features in other ethical views, including Rule Utilitarianism. It also discusses some of the issues facing a theory of pattern-based reasons, especially the difficulty of giving a satisfactory account of the 'eligibility' of patterns—that is, an account of which patterns generate reasons, and which do not. In this chapter I depart significantly from some claims about eligibility that I have defended in previous work.²⁰

Chapter 6 discusses utilitarian attempts to account for moral rights. I claim that utilitarians should try to explain moral rights, but that existing attempts are open to serious objections. I then try to show that we can explain moral rights using the concept of pattern-based reasons. The basic idea is that there exists a valuable practice which is such that playing one's part in it amounts to respecting others' moral rights, and that we have sufficient (pattern-based) reasons to participate in this practice. The chapter also discusses whether rights are too contingent according to this picture.

Chapter 7 begins by outlining a utilitarian account of justice. I argue that utilitarians can construct a plausible theory of justice, as respect for moral rights. Since moral rights include rights that depend for their existence on positive laws, justice will, on this view, sometimes consist in seeing to it that people get what they are legally entitled to. This implication conflicts with many current theories of distributive justice, including egalitarian theories. However, later in the chapter I argue that utilitarians can explain many egalitarian intuitions, even though they will treat some egalitarian aims as matters of good social policy rather than as requirements of justice.

²⁰ In Woodard 2008a and 2017 I rejected what I called the 'willingness requirement'. As I explain in Chapter 5 below, I now accept a form of this requirement.

One of the six main objections identified in Chapter 2 is that utilitarianism fails to recognize the distinctiveness of political issues from moral issues. In Chapter 8 I argue that utilitarians can account for the normative significance of legitimacy. That is, they can account for intuitions to the effect that the fact that a decision was taken legitimately gives us some reason to comply with it even if it was the wrong decision on independent grounds. Furthermore, I argue that utilitarians can also give qualified support to democratic procedures. The normative significance of legitimacy and democracy enables utilitarians to explain at least some of the distinctiveness of political issues.

Chapter 9 attempts to discharge the debt incurred in Chapter 3, of explaining praiseworthiness, blameworthiness, and the nature of good deliberation. I argue that we do not need a perspectival concept of reasons, or a perspectival concept of rightness, in order to explain these things. A good deliberator is one who possesses psychological characteristics that generally result in right action. An action or character trait is praiseworthy or blameworthy if it is picked out as such by a practice of praising and blaming in which we have reason to participate. Putting these two ideas together, we can define virtues as character traits that are both features of good deliberators and also praiseworthy. The result is a picture of virtuous agents, good decision-making, and praise and blame in which the concept of perspectival reasons plays no essential role.

Chapter 3–9 present the outline of a utilitarian theory that spans topics in moral and political philosophy. I do not claim that it is a complete theory of these topics; indeed, I try to emphasize the gaps, to indicate areas for further development. Nevertheless, I claim in Chapter 10 that the outline, such as it is, enables us to give plausible answers to the objections identified in Chapter 2. If that is correct, then utilitarianism is worth taking seriously.

2

Six Objections

Utilitarianism would not be worth taking seriously if there were one or more clearly fatal objections to it. It might then have some interest to the historian of ideas, but its interest to those engaged in normative ethics or political philosophy would be limited to illustrating how not to proceed. However, I do not think that there is a clearly fatal objection to utilitarianism; at least, not one that has already been discovered.

I shall try to show this by discussing a number of influential objections to utilitarianism. This chapter assembles the six main objections to be discussed later. The list is, of course, selective. Utilitarianism has been subject to much critical discussion, and I can discuss only a small portion of it. Nevertheless, the selection had better not dodge any of the main kinds of objection. My case for taking utilitarianism seriously depends on showing that it can respond plausibly to the main objections.

In considering the objections we should remember that choice of theory in normative ethics and political philosophy is ultimately a comparative exercise. Some problems for a given theory also afflict the alternatives to it.¹ Not all problems are like that, of course, and so we should try to note as we go along the objections that are particularly pressing for utilitarianism. We should remember also that we are defining utilitarianism as a family of views, and so we should note instances in which an objection applies only to some members of the family, or more forcefully to some than to others.

None of the six kinds of objection discussed in this chapter is easily answered. All of them pose significant challenges to utilitarianism. Nevertheless, as I hope to show later in the book, none of them is clearly fatal.

¹ This is true of two important problems that I will largely leave aside: population ethics, and the ethical status of non-humans. Though utilitarianism faces difficult problems in these areas, so do its main rivals. On consequentialism and population ethics, see Parfit 1987: pt. 4 and Mulgan 2006. On utilitarianism and the ethical status of non-human animals, see Višák 2013 and Višák and Garner 2016.

2.1 Pig Philosophy

The essayist Thomas Carlyle nicely encapsulated a common criticism of utilitarianism when he referred to it as a ‘pig philosophy’. The root of Carlyle’s criticism appears to have been the idea that concern with happiness crowds out nobility of action:

Unfortunate creatures! . . . Only ignoble, expensive and unfruitful things can you now do; nobleness has vanished from the sphere where you live. The way of it is lost, lost; the possibility of it has become incredible. We must try to do without it, I am told.—Well; rejoice in your upholsteries and cookeries, then, if so be they will make you ‘happy’.²

Carlyle’s specific criticism is aimed at hedonist forms of utilitarianism, which claim that goodness consists in happiness, understood as the balance of pleasure over pain. On some crude views of happiness or pleasure, the claim that pursuit of happiness crowds out noble pursuits may seem correct. If happiness is a matter only of upholstery and cookery, then pursuit of it may be ignoble. But as John Stuart Mill and others have pointed out, this criticism seems to suggest that the critic has an unduly narrow conception of happiness. It is not obvious that it identifies a problem with hedonism itself.³

Carlyle’s criticism—that utilitarianism is a pig philosophy—appears to be directed at crude forms of hedonist utilitarianism. This is a very narrow target, since it leaves out more sophisticated forms of hedonism as well as non-hedonist forms of utilitarianism.⁴ However, we can broaden Carlyle’s point, and interpret it as the claim that utilitarianism rests on a mistaken theory of value. Taken in that broader way the criticism is not easily dismissed, and it remains a recurrent theme of criticisms of utilitarianism.⁵ Moreover, given utilitarianism’s ambition to explain ethical phenomena in terms of their relations to value, this kind of criticism is potentially very damaging. Utilitarianism cannot offer the correct explanation of rightness, virtue, or other ethical phenomena if it tries to explain them all in terms of value, but it embodies a false understanding of value.

² Carlyle 1850: 286. Nietzsche also criticizes utilitarianism for being ignoble, like other forms of ‘slave morality’. See for example Nietzsche 2002/1886: sections 225, 228, and 260; see also Anomaly 2005.

³ Mill 1998/1861: ch. 2.

⁴ See Feldman 2004: 38–40, 72–8, and 117–22 for a discussion of some options open to hedonists.

⁵ See for example Moore 1993/1903: chs. 3 and 6, Rawls 1999a: 19–30, Sen 1979, and Kraut 2007: ch.1.

In fact, utilitarian theories each offer two theories of value, which should be distinguished. One, the most fundamental, is an account of *well-being*—understood as synonymous with ‘utility’ and ‘welfare’. Well-being is the noninstrumental value of a life for the person whose life it is, so a theory of well-being is a theory of this kind of value. In addition, utilitarianism contains a theory of the value of *outcomes*, according to which the value of an outcome is equal to the sum of the well-being it contains.⁶ This theory is the conjunction of two claims. *Welfarism* claims that all and only well-being has noninstrumental value. All well-being has noninstrumental value, and nothing else does. *Sum-ranking* is a further claim, to the effect that the value of an outcome is the sum of the goods (and bads) existing in that outcome.⁷ Thus there are three main targets for critics of utilitarians’ claims about value: their claims about the nature of well-being; their commitment to welfarism; and their commitment to sum-ranking.

So far as the account of the nature of well-being goes, one might assume that the utilitarian is on strong ground. Though doubts can of course be raised about any particular theory of well-being, utilitarianism is not committed to any particular theory of well-being. For example, if a utilitarian were to become convinced that hedonism is not the correct account of well-being, she could simply renounce hedonism. Going further, it is tempting to assume that utilitarians can adopt whichever theory of well-being turns out, all things considered, to be most credible.

However, this response is too relaxed. The use to which utilitarians wish to put the theory of well-being places some constraints on which theory they can adopt. As consequentialists, they wish to use the theory of well-being to explain the rightness of actions. This means that they cannot without circularity adopt a theory of well-being according to which, for example, well-being consists in *pleasure derived permissibly*. Yet some have wanted to introduce just such a restriction into their theory of well-being, to capture the judgement that pleasures derived through cruelty or other morally wrong action do not add to the subject’s well-being.⁸

⁶ Consequentialists, including utilitarians, define ‘the outcome of *X*’ broadly, to include everything that would happen were *X* realized. In the case of actions, for example, they include the performance of an action amongst its consequences or outcome. For example, see Broome 1991: 3–4 and Portmore 2011: 34.

⁷ The term and the idea are from Sen 1979: 468, though he defines it slightly differently.

⁸ Rawls 1999a: 27 contrasts utilitarianism with his own theory of justice, according to which, ‘[t]he principles of right, and so of justice, put limits on which satisfactions have value.’ Scanlon 1998: ch. 3 also discusses constraints placed by moral theories on theories of well-being.

Second, since utilitarians are welfarists they cannot appeal to any theory of well-being that posits the existence of noninstrumental value that is not itself well-being or a constituent of well-being. Now, this constraint is tricky to interpret, because well-being could be internally complex. The best theory of well-being might claim that it consists, for example, in possession of subordinate goods such as friendship and pleasure. This is consistent with welfarism, so long as the value of these subordinate goods is entirely a matter of their contribution to well-being. But that is itself a substantive constraint on which theory of well-being can be adopted by the utilitarian. She could not, for example, adopt a theory according to which one of the constituents of well-being is proper appreciation of the well-being-independent value of beauty. And as Roger Crisp has pointed out, it is not enough simply to say that the constituents of well-being, though apparently themselves kinds of value, do not have any value independent of well-being. For there will be a question about why, exactly, they do not have any independent value, and it may be difficult to answer this question in a way that is convincing and consistent with welfarism.⁹

For these two reasons, utilitarians must proceed cautiously in adopting a theory of well-being. There is some room for manoeuvre, perhaps, in accepting a degree of circularity by downgrading their explanatory ambitions, or in postulating internal complexity in well-being. But utilitarians should not blithely assume that, whatever the best theory of well-being turns out to be, it can be plugged into their broader theory.

Now consider the utilitarian theory of the value of outcomes. As welfarists, utilitarians claim that all and only well-being has noninstrumental value. Obviously, this claim can be challenged from two directions. First, does *all* well-being have noninstrumental value? Sadists desire that others suffer, and derive pleasure from that suffering. One question is whether the satisfaction of this desire, or the pleasure that it brings to the sadist, contributes to the sadist's well-being. As we have just noted, the utilitarian's commitment to consequentialism narrows her options for denying this. Suppose, then, that she accepts it. The next question is whether the sadist's well-being makes the outcome better. The utilitarian's commitment to welfarism forces her to say that it does. Of course, the utilitarian is not likely to say that, all things considered, it is good that the sadist's desire is satisfied. But she does seem committed to saying that, holding the level of suffering

⁹ Crisp 1998: 12 raises this question about Mill's qualitative hedonism. See also Crisp 2000, McNaughton and Rawling 2001, and Crisp 2001.

fixed, the fact that the sadist gets some additional well-being from the suffering makes the outcome better.

Second, welfarism can be challenged from the opposite direction. Does *only* well-being have noninstrumental value? Many think that other things, such as beauty, knowledge, virtue, and justice, also have value.¹⁰ As welfarists, utilitarians must claim about each of these candidate additional values either that it is not valuable, or more plausibly that its value is entirely dependent on its contribution to well-being. The simplest kind of contribution is *causal*: beauty, for example, might be said to be valuable insofar as it causes pleasure, and so makes an instrumental contribution to well-being. But as we noted above, it is also possible to claim that well-being consists in the possession of subordinate goods. We might then claim that these other candidate values make a *constitutive* (not merely a causal) contribution to well-being. Now, in fact, the claim that these other things make a causal or constitutive contribution to well-being is very plausible. This seems true of beauty, knowledge, virtue, and justice, for example. So the crucial question is whether this exhausts their value. Like other welfarists, utilitarians must claim that it does.

The third target for critics of utilitarianism's claims about value is its commitment to sum-ranking. Welfarism tells us that the goodness of an outcome depends only on the well-being it contains; sum-ranking specifies the precise form of this dependency. It claims that the goodness of the outcome equals the total of the well-being it contains, so that one outcome is better than another if and only if it contains more well-being. Importantly, not all welfarist theories also endorse sum-ranking. One important alternative is to rank outcomes by their average level of well-being. This ranking coincides with sum-ranking in all cases except those in which the population varies, so the difference is especially important with regard to population ethics.¹¹ Another important alternative is to rank outcomes solely according to the well-being of the worst-off. Alternatively, one could give extra weight to the levels of well-being of the worse-off, while giving some weight to the levels of the better-off.¹²

These alternatives to sum-ranking deserve serious consideration. They can be used to generate a set of close cousins of utilitarianism as I have

¹⁰ Moore 1993/1903: 135–6 claims that beauty is good in itself. Ross 2002/1930: 140 claims that 'four things . . . seem to be intrinsically good—virtue, pleasure, the allocation of pleasure to the virtuous, and knowledge'. The distribution of pleasure according to virtue is one conception of justice.

¹¹ See Parfit 1987: pt. 4.

¹² See Parfit 2000: 103–5 and Hooker 2000: 55–65.

defined it, which share its commitment to consequentialism and welfarism but not its commitment to sum-ranking. We will return to some of them in Section 2.4 below. As we will see there, they speak to powerful intuitions to the effect that the way in which utilitarianism aggregates the well-being levels of each individual is too simple.

Thus utilitarianism is open to serious challenges to its claims about well-being and about the value of outcomes. Suitably refined, Carlyle's objection may still have teeth. Moreover, objections to welfarism and sum-ranking threaten to undermine every form of utilitarianism, since they are all committed to both of these claims.¹³

2.2 Abhorrent Actions

A second venerable objection focuses on utilitarian claims about right actions. In particular, this objection claims that utilitarianism wrongly implies that some *abhorrent actions* are morally permissible. These are actions that, the critic supposes, our strong convictions condemn. The point of the criticism is to show that these convictions conflict with utilitarianism, and thereby to cast doubt on utilitarianism.

Bishop Butler made a version of this objection when he pointed out that robbery could, on some occasion, promote utility, since the loss to the victim could be outweighed by a gain to the beneficiary—'yet such an action would surely be vitious'.¹⁴ Sir David Ross made a version of it when he pointed out that Act Utilitarians should favour breaking a promise whenever doing so would lead to even slightly more utility than would keeping it.¹⁵ H. J. McCloskey made another version when he pointed out that utilitarians should favour punishing the innocent, or punishing someone out of proportion to her crime, when doing so maximizes utility.¹⁶

The general form of the objection from abhorrent actions is to identify a kind of action, *K*, and to claim, first, that utilitarianism sometimes permits actions of this kind, and second, that on at least some of these occasions,

¹³ Sen 1979: 468.

¹⁴ Butler 2006b/1736: 312. Ewing 1948: 109 expressed the objection by asking: 'Is it really right... to further the greatest good by all means in our power including, if necessary, the most caddish and abominable, and, if not, how can utilitarianism be defensible?'

¹⁵ Ross 2002/1930: 34–5. Carson 2005 criticizes Ross's reliance on intuitions about promise-keeping.

¹⁶ See McCloskey 1967. Hurka 2014: 166–7 traces the history of this objection, claiming that it was first made in a paper by Ewing in 1927.

actions of this kind are clearly morally wrong. This sort of objection applies most straightforwardly to Act Utilitarianism in particular. Act Utilitarianism claims, roughly, that an action is morally permitted so long as no alternative to it in the circumstances would generate more utility. Thus Act Utilitarianism ties the rightness of an action to its consequences, rather than its kind. The kind of action could be robbery, or promise-breaking, or unjust punishment—or, in fact, any other kind, no matter how gruesome. Dostoevsky goes considerably further than Ross or Butler in this respect, in having his character Alyosha contemplate a case in which torturing and killing a small child would maximize utility.¹⁷

However, ramping up the horror risks distorting the criticism. It suggests that the underlying issue is whether actions of the horrific kind are *ever* permitted, in any circumstances. The accusation would then be that Act Utilitarianism is objectionable because it does not endorse absolute prohibitions of these kinds of actions: that is, it does not claim that, in every possible circumstance, such actions are morally wrong. For some critics of utilitarianism, that may indeed be the salient issue.¹⁸ But many other critics themselves reject absolute prohibitions, and so on this construal of the objection they are on the same side as Act Utilitarianism. Robert Nozick, for example, criticizes utilitarianism for its failure to understand the moral significance of individuals' rights, but stops short of endorsing an absolute prohibition against violating rights.¹⁹ Similarly, Ross did not claim that it is always wrong to break a promise. These moderate critics of utilitarianism must have a subtler point in mind.

The issue seems to be what kind of moral significance, if any, attaches to an action's kind. Theories that endorse absolute prohibitions attribute a very strong sort of significance: they claim that being of the prohibited kind is sufficient for wrongness. At the other extreme, Act Utilitarianism attributes no independent moral significance to an action's kind: it claims that an action's kind matters only to the extent that it affects its consequences (as when, for example, the fact that an action was a lie upsets someone). Between these extremes there is a good deal of logical space. In particular, there is room for views according to which an action's being of a certain kind

¹⁷ Dostoevsky 1993/1880: 282.

¹⁸ For example, see Anscombe 1958: 9–19. Kant also endorsed exceptionless duties, of course. On the exceptionlessness of the duty not to lie, for example, see Kant 1996b/1797: 612–13.

¹⁹ Nozick writes: '[t]he question of whether these side constraints are absolute, or whether they may be violated in order to avoid catastrophic moral horror, and if the latter, what the resulting structure might look like, is one I hope largely to avoid' 1974: 30n.

is sufficient for there to be a *reason* for or against it—what we might call ‘general reasons’.

This seems to be what the moderate critics have in mind. And for this purpose Ross’s case is more deftly chosen than Dostoevsky’s. Dostoevsky’s horrific action with dramatic consequences is salient if we are considering whether there are absolute prohibitions against some kinds of action. In contrast, Ross’s example draws attention to the fact that Act Utilitarians recognize no reason to keep a promise that is independent of the consequences of doing so—none, that is, that attaches to the kind of action, *keeping promises*, itself. Since there is no reason of this sort to keep a promise, the slightest benefit suffices to make breaking a promise morally right. Ross’s focus on small benefits is thus more apt for the purpose of drawing attention to the issue of whether reasons attach to kinds of action.

According to Ross, there are some circumstances in which breaking a promise is wrong even though it has the best outcome. In modern parlance, this is to accept an agent-relative constraint—or ‘constraint’ for short.²⁰ Constraints imply that it is sometimes wrong to bring about the outcome that is best from an impartial point of view, such as the point of view specified by the utilitarian theory of the value of outcomes. Constraints result when the reason associated with a kind of action is sufficiently weighty to make it wrong, at least sometimes, to bring about the best outcome by performing or failing to perform an action of that kind. Though their name perhaps suggests otherwise, they can be the result of reasons *to* perform actions of a certain kind, as well as reasons *against* performing actions of a certain kind. Thus there could be a constraint against killing, resulting from a general reason against killing. But there could also be a constraint requiring generosity—even when this stands in the way of bringing about the best outcome—resulting from a general reason to be generous.

The objection from abhorrent actions thus cites kinds of action that, it is supposed, we think are governed by constraints. One particularly important example of this sort of criticism is the claim that Act Utilitarianism is unacceptable because it requires agents to violate rights whenever doing so is necessary to bring about the best outcome.²¹ Another important example is the claim that Act Utilitarianism is unacceptable because it does not recognize the moral significance of so-called ‘special obligations’.²² These are obligations

²⁰ Kagan 1989: 4, 24–32. Scheffler 1994: 2–3 calls constraints ‘agent-centred restrictions’.

²¹ For example, see Rawls 1999a: 137–9, Nozick 1974: ch. 3, Sen 1982, Lyons 1994: ch. 6.

²² Jeske 1998.

owed to specific others in virtue of one's relationship to them (for example, as parent) or in virtue of one's undertakings to them (for example, as signatory to a contract). Both rights and special obligations might be said to be governed by constraints. Both also offer ways of making sense of the intuitively important distinction between *acting wrongly* and *wronging someone*.²³ To wrong someone is to commit a wrong that is specifically directed at her, and an obvious way to explain that directedness is to say that it involves violation of one or more of her rights or one or more special obligations owed to her.

In its broadest form, then, the objection from abhorrent actions is an objection to the effect that utilitarian theories fail to explain the moral constraints on our actions. This objection applies most straightforwardly to Act Utilitarianism, since it does not recognize any constraints at all. Other forms of utilitarianism, such as Rule Utilitarianism and Motive Utilitarianism, do recognize constraints, and general reasons attaching to kinds of action. But it does not follow that they are immune to the objection from abhorrent actions. These other forms of utilitarianism may not endorse just the set of constraints that, intuitively, we think exist.²⁴

2.3 Demandingness

The objection from abhorrent actions targeted utilitarian claims about what is morally permitted. Another very influential objection targets utilitarian claims about what is morally required. According to this objection, utilitarianism should be rejected because it is too demanding of agents: it requires too much of them.

Once again, the objection applies most straightforwardly to Act Utilitarianism. However, there is a danger in discussing it mainly as it applies to Act Utilitarianism: we may neglect the fact that it arises for many or most other moral theories as well. Kantianism and virtue ethics are subject to forms of this objection, as are alternatives to Act Utilitarianism within the family of utilitarian views.²⁵ However, the objection was made initially against Act Utilitarianism, and it is easiest to see how it arises there.

²³ Nelson 2015 claims that utilitarianism cannot make sense of this distinction.

²⁴ For example, McCloskey 1967: 92 claims that, like Act Utilitarianism, Rule Utilitarianism is incapable of explaining reasons not to punish unjustly.

²⁵ On the demandingness of Kantian ethics see van Ackeren and Sticker 2015. On virtue ethics and demandingness see Swanton 2003: ch. 9. On the demandingness of Rule Consequentialism see Carson 1991, Hooker 2000: ch. 8, Mulgan 2001: ch. 3, and Hooker 2009.

Consider what Act Utilitarianism requires of each of us as agents. According to standard forms of Act Utilitarianism, any action that fails to maximize utility is morally wrong; each of us is morally required, at all times, to maximize (actual or expected) utility. The problem is then the mismatch between the things we could do to promote utility, on one hand, and the things we typically do and typically take to be morally innocent, on the other. Most of us do much less than we could to ameliorate the vast amount of poverty and suffering in the world. This poverty and suffering has huge disutility, and each of us can do things to help. This could be a matter of volunteering or working for a charity, though it may be more effective for many people to ‘work to give’—that is, to pursue some lucrative career in order to be able to give more to utility-promoting causes.²⁶ Act Utilitarianism appears to require each of us to devote our lives to such causes, and to leave no room for activities that fail to maximize utility but which we normally think are morally innocent, or even admirable, such as cultivating friendships, pursuing worthwhile personal projects, or enjoyably loafing around.

Not only that: Act Utilitarianism requires that every opportunity to maximize utility be taken. So if you are already working to give, but could give a little more, it requires you to do that. If you could also bring a little more joy to others while you are at it, it requires you to do that as well. If you are telling a moderately amusing joke but could have told a devastatingly funny one, then it says you are acting wrongly. And although the demands of Act Utilitarianism fall most obviously on members of affluent countries, they also fall on the global poor themselves. For everyone, no matter how impoverished, has some opportunities to promote utility, and Act Utilitarianism requires that they be used to the greatest possible effect.

According to the demandingness objection, these demands are excessive: not just large, but too large. This point can be put in different ways, each associated with a different form of the objection. One issue is whether to focus on what Act Utilitarianism requires agents to *do*, or instead on what it requires them to *care about*.²⁷ If we focus on actions, we may then characterize the demands in terms of cost to the agent’s self-interest, or in terms of disruption to her broader personal projects or—less grandly—simply her desires.²⁸ The size of costs, and the idea of an *excessive* cost, could be spelled

²⁶ See MacAskill 2014.

²⁷ Brad Hooker 2009: 155 draws this useful distinction very clearly.

²⁸ See Kagan 1989: 233–51. Chappell 2017 proposes understanding demandingness in terms of difficulty rather than cost.

out in terms of one-off or cumulative demands.²⁹ Alternatively, we could express the objection by saying that Act Utilitarianism wrongly fails to distinguish between actions that are morally required, and those that go admirably beyond the call of duty—the ‘supererogatory’.³⁰ If instead we focus on what Act Utilitarianism requires agents to care about, we could express the objection by saying that it requires agents to have an alienating degree of impartial concern, or that it undermines their integrity by requiring them to subordinate their personal projects to the impersonal goal of maximizing utility.³¹

Though there are important differences between these different ways of developing the objection, it makes sense for present purposes to gather them under the umbrella term ‘the demandingness objection’. All of them point to one or more aspects of the mismatch mentioned above, between what Act Utilitarianism appears to require of agents, and that subset of what we in fact do and care about that we take to be morally permissible.

It is hard to deny that there is such a mismatch.³² But as soon as we reflect on how to respond to it we run into difficult methodological issues. One issue is whether to give priority to our convictions to the effect that ordinary non-utility-maximizing activities are morally permissible, or to give priority to the arguments for Act Utilitarianism. Viewed in one light, those convictions are just the sort that we should not trust, since they serve our vested interest in carrying on our lives as normal.³³ Viewed in another light, these convictions express some of our deepest commitments—too deep to be dislodged by an abstract argument for a moral theory.³⁴

A second issue is that, as David Sobel has argued, the demandingness objection presupposes some doctrines that no Act Utilitarian would accept, such as the greater moral importance, for the agent’s reasons, of costs to her than costs of the same size to those she could help. Without some such

²⁹ Cullity writes: ‘it seems that I could almost always be doing something equivalent to saving a life at tiny personal cost. Although the cost of each such action might be trivial, there are so many desperately needy people in the world that iterating this requirement would impose a total cost that is very high’ 2009: 13.

³⁰ On utilitarianism and supererogation see Vessel 2010.

³¹ See Williams 1973: 93–118, Railton 1984, and Brink 1986.

³² See Mulgan 2001: 31–7. Bruder and Tanyi 2014 presents experimental evidence on the degree to which intuitions about demandingness are shared.

³³ Rawls 1999a: 42 excludes from the category of considered judgements those made ‘when we stand to gain one way or the other’. Many utilitarians have sought to defend their view in part by expressing scepticism about appeals to intuition. See for example Brandt 1979: 16–23, Hare 1952: 64, 74–8, and 1989: ch. 10, and Singer 2005.

³⁴ Mulgan 2001: 26–31. See also Scheffler 1992: ch. 7.

doctrine, we would have to say that any moral view that does not require each agent to help others when doing so maximizes utility thereby permits those who go unaided to bear greater costs than Act Utilitarianism imposes on the agent. It thus demands more of them than Act Utilitarianism demands of agents. Thus, the demandingness objection is not a standalone criticism.³⁵ This seems correct, but it does not follow that it is not a criticism at all. The best discussions of demandingness recognize that the issues it raises involve conflict between our moral theories and a complex bundle of ordinary moral convictions—not just a single, independent intuition about demandingness.³⁶

If we are minded to seek a less demanding theory than Act Utilitarianism there are two main strategies to consider. The first seeks a picture of morality according to which it is less demanding. The way to do this is to go for a moral theory that endorses so-called agent-relative options—‘options’ for short. An option is a moral permission for the agent to favour her own interests, projects, or desires, up to some limit.³⁷ Whereas constraints forbid the agent from maximizing the impartial good, options permit her to maximize the impartial good but also permit her not to do so, within some limits. In cases where we are worried about intuitively excessive demands on the agent we do not typically think that she is forbidden from maximizing the impartial good; instead we think that morality does not require her to do so. Options capture exactly this thought.

The second strategy is to reduce the practical significance of the demands of morality by reducing the status of morality itself. If moral reasons do not always override other reasons, then agents may sometimes have most reason to pursue their own interests or projects even when morality demands that they set these aside. This strategy seeks to accommodate worries about

³⁵ Sobel concludes: ‘Concerns about demandingness might be helpful in locating counter-intuitive consequences of Consequentialism, but to vindicate such thoughts we must vindicate other concerns about Consequentialism, concerns such as that there is an important moral distinction between causing and allowing’ 2007: 17.

³⁶ See for example Kagan 1989 and Mulgan 2001.

³⁷ Scheffler 1994 proposes a theory (originally published in 1982) that includes options (which he calls ‘agent-centred prerogatives’), but not constraints (which he calls ‘agent-centred restrictions’). Kagan 1989: 19–24 criticizes this combination, noting that it appears to countenance an option to do harm as well as an option to allow harm. Scheffler 1994: Appendix 3 replies. Note that the term ‘option’ has the disadvantage of being widely used to refer to a different idea, as well: in this other sense, ‘option’ refers to the range of actions an agent is able to perform. Despite this, ‘option’ is more widely used than ‘prerogative’, and I will adopt it in what follows.

demandingness by letting morality's demands be overridden, rather than by reducing morality's demands.³⁸

Both strategies are open to utilitarians. On the face of it Rule Utilitarianism, for example, is likely to be much less demanding than Act Utilitarianism, since it conceives of promoting the good as a joint enterprise. This seems to share the burdens more widely, and to support the existence of options. For different reasons, Motive Utilitarians can also accommodate options. But as with the discussion of constraints, the mere possibility of accommodating options does not guarantee that these theories will endorse just the options that we think, on reflection, exist. Nor does the strategy of allowing moral demands to be overridden come without theoretical costs. The demandingness objection thus remains an important challenge to any form of utilitarianism.

2.4 Separateness of Persons

So far we have considered objections to what utilitarianism says about value, what it says agents are permitted to do, and what it says agents are required to do. The next objection touches on all of these issues, but aims to diagnose why utilitarians (allegedly) get these things wrong. According to the objection from 'the separateness of persons' they do so because the way they approach moral issues fails to take proper account of the distinctness of individuals, each with their own life to live.

This objection is due to John Rawls. Rawls claimed that utilitarianism treats conflicts of interest among people on the model of rational prudence. Just as the rationally prudent person judges a loss at one time to be justified by a larger gain at another, utilitarians judge a loss to one person to be justified by a larger gain to one or more other people. But, according to Rawls, this extension of the model of rational prudence to cover cases of conflict of interest between persons is illegitimate, and amounts to failure to 'take seriously the distinction between persons'.³⁹

Rawls's claim is alluring, because it promises to identify a deep flaw in utilitarianism. But it is also very slippery, because it is not at all clear exactly

³⁸ Significant discussions of this idea include Sidgwick 1907: 496–509, Brink 1986: 434–38, Scheffler 1992, Sobel 2007: 14–16, Portmore 2011: ch. 5, and Parfit 2011: ch. 6.

³⁹ Rawls 1999a: 24. Rawls's claim that utilitarianism treats society as if it were a single person was anticipated in Myrdal 1953: 54. Myrdal added, as if in mitigation, that utilitarianism had spared economic theory from the 'horrors' of German political philosophy.

what the mistake is supposed to be. To appreciate the genuine issues it raises, we have to strip away some possible misconceptions.

The first misconception is to think that Rawls's point is that utilitarianism somehow fails to notice or attach any significance to the fact that persons are distinct from one another. That cannot be right, and cannot be what Rawls meant. Utilitarianism is an individualistic theory, in the sense that it claims that the goodness of outcomes is entirely a matter of their goodness for the individuals they contain. Total utility is the sum of the utilities of these individuals. Moreover, utilitarians are typically highly sensitive to the fact that what suits one person will not suit another. Thus they can hardly be accused of overlooking the fact that persons are distinct from one another.

Instead, the objection has to be that the way in which utilitarians respond to facts about what is good or bad for each person fails to take seriously the distinctness of persons. But what exactly in their response to these facts is supposed to be mistaken?

The thought could be that utilitarians wrongly treat different persons' interests as fungible or interchangeable. Other things equal, utilitarians treat an outcome in which Sally has 10 and Sabiha has 5 as being exactly as good as an outcome in which Sabiha has 10 and Sally has 5. This can give the impression that utilitarians treat persons merely as locations or 'receptacles' of value. On this interpretation, they should be completely indifferent between these outcomes. But as Richard Yetter Chappell has argued, this is a misinterpretation of utilitarianism. Utilitarians are not committed to the claim that there is one ultimate good, total utility. Instead they can claim that each person's well-being is a *separate* ultimate good, and thus a separate source of reasons. On this basis they should feel ambivalent about the two outcomes involving Sally and Sabiha, not indifferent between them. Each outcome is better with respect to one of the distinct goods that ground reasons, and worse with respect to the other.⁴⁰

The real issue is aggregation of interests. Their commitment to sum-ranking means that utilitarians aggregate interests in a very simple way: they add them up. The value of an outcome equals the sum of its values for all of the *subjects* involved in it, at all the *times* it includes. As I have defined utilitarianism, all utilitarian theories share this commitment. This means

⁴⁰ Chappell 2015: 324–9. It is as if on a single day the *Mona Lisa* is destroyed but we also find an equally good and previously undiscovered work by da Vinci. No art lover should feel indifferent about this development, though ambivalence seems appropriate precisely because the value of each work is the same.

that they are committed to claims about the value of outcomes that may seem dubious or worse.⁴¹

Start by considering what sum-ranking says about the value of a single life: the value of a subject's life is equal to the sum of her levels of well-being at each time. How well-being is distributed across these times does not matter, except as it affects the total. Thus, a life in which things start poorly then steadily improve until death is equally good as one that has the exact opposite trajectory, provided the sum total of well-being in each life is the same.⁴² A life in which a fixed sum of well-being is spread very thinly over thousands of years is equally good as one in which that sum is concentrated within eighty years, or concentrated still further in one year.⁴³ A life in which well-being is constant at all times is equally as good as one in which there are dramatic peaks and troughs, if the total in each is the same.

It is hard to think clearly about these cases. Considering the improving and declining lives, it is hard not to be distracted by imagining the benefit of looking forward with optimism that would characterize one, and the harm of looking backward wistfully that would characterize the other. But such harms and benefits must be included in the totals, which are stipulated to be the same. Considering the very extended life in which well-being is spread very thinly, it is hard to grasp the positive value of the well-being at each moment, though this feature is stipulated to be present. Considering the constant and varying lives, it is hard not to add an extra bonus for excitement or a further penalty for boredom.

Now consider sum-ranking as it applies across lives. The value of an outcome is the sum total of the well-being of all the subjects it contains, and the well-being of each subject is the sum total of her well-being at all times. According to sum-ranking, how well-being is distributed across these subjects and times does not matter, except as it affects the total. Nor, when we compare the value of outcomes, does it matter whether they each contain the same subjects, or the same number of subjects. Each outcome's value depends only on the total amount of well-being it contains.

This too seems very implausible to many. Critics point to a wide range of factors to which it denies fundamental significance, and claim that one or

⁴¹ Frankel Paul et al. 2008 collects papers discussing issues of aggregation from a variety of perspectives.

⁴² On the significance of the shape of a life see Velleman 1991, Feldman 2004: ch. 6, and Glasgow 2013.

⁴³ This claim gives rise to an intrapersonal version of Parfit's 'repugnant conclusion' 1987: ch. 17. See Temkin 2012: ch. 4 and Hirose 2015: 96–8.

more of them is fundamentally significant. For example, sum-ranking ignores inequality and other features of the distribution of a fixed total of well-being across a fixed population of subjects. The outcome (50, 50) is no better than the outcome (99, 1), or the outcome (200, -100).⁴⁴ Sum-ranking also ignores the identity of subjects: (10, 90) is exactly as good as (90, 10). Further, it ignores the number of subjects: (50, 50) is exactly as good as (25, 25, 25, 25).⁴⁵ Finally, it ignores the deservingness of the subjects: a world in which the saints are happy and the sinners are miserable is no better than a world in which the sinners are happy and the saints are miserable, if the totals of utility are the same.⁴⁶

Again, these claims are hard to assess. It is hard, as before, to set aside strictly extraneous thoughts, such as imagining being resentful of inequality, or imagining the likely bad effects of sinners doing better than saints. Admittedly we can adopt useful devices to help guard against this, such as imagining that the different subjects are isolated from one another, so that resentment and incentive effects cannot take hold.⁴⁷ Still, it is hard to be sure that our judgements about the value of these outcomes are not distorted by such factors. This is especially true when the examples involve very large numbers of people, or very small but positive amounts of well-being.⁴⁸

It is also important to distinguish claims about value from claims about what any agent ought to do, or has reason to do. So far, we have been considering sum-ranking taken by itself. By itself, it yields only claims about value. It does not yield claims about reasons or rightness. With respect to purely axiological claims, utilitarianism as I have defined it has no room for manoeuvre, since sum-ranking is one of its defining features. So utilitarians as I have defined them must simply accept the implications of sum-ranking.

⁴⁴ Each number represents the interpersonally comparable level of utility of one person. Roemer 1996: 160 describes this implication of utilitarianism as 'disturbing'. See also Nozick 1974: 41 on the possibility of a 'utility monster'.

⁴⁵ This fact gives rise to the so-called 'repugnant conclusion'. See Parfit 1987: ch. 17 and Huemer 2008.

⁴⁶ For the claim that it is better noninstrumentally that the virtuous do well than that the vicious do equally well, see Ross 2002/1930: 35, 138, Temkin 2000: 138–40, and Feldman 2004: ch. 9. Kagan 2012 is an extended discussion of the possible axiological significance of desert.

⁴⁷ This device is used in Parfit 2000: 87.

⁴⁸ Broome 2004: 58 writes: 'We have no grounds for thinking that our intuitions about very large numbers are reliable. On the contrary, we have good grounds for mistrusting them. The evidence is that they are often wrong, because our imagination is not able to grasp just how big numbers can be. For example, many people's intuition tells them that the process of natural selection, however many billions of years it continued for, could not lead from primordial slime to creatures with intelligence and consciousness. But they are wrong. Four billion years will do it.' See also Cowie 2017.

If they find those implications counter-intuitive, they can at most point to some of the difficulties involved in these particular judgements, as I have just done, and point to the counter-intuitive implications of rival theories of the value of outcomes, as I have not done.

When it comes to what any agent ought to do or has reason to do, things are very different. Among all possible utilitarian views, only standard forms of Act Utilitarianism are committed to the claim that agents ought always to bring about the best possible outcome and have most reason to do so. Other versions of utilitarianism can deploy whatever additional complexity their theory affords to try to explain our judgements about when it is right to bring about the best outcome and when it is not. In particular, they will hope to explain a sufficient number of our convictions about when it is right to impose costs on someone for the sake of increasing total utility.

The success of Rawls's attempt to diagnose a deep fault in utilitarianism depends greatly on whether we accept sum-ranking as a claim about the value of outcomes. I have claimed that assessing sum-ranking is harder than it may seem. If we reject it, we should do so in favour of a better theory of the value of outcomes, and it is not clear what that better theory is.⁴⁹ On the other hand, if we accept sum-ranking, the real issues raised by Rawls's discussion concern when and why costs should be imposed on people for the sake of making the outcome better. Different utilitarian theories will give very different answers to these questions, and it is quite possible that one or more of them will give satisfactory answers—including explaining why, for example, it seems wrong to impose great costs on someone who is badly off for the sake of small benefits to very many better off people.

2.5 Politics

So far we have been considering utilitarianism exclusively as a moral theory. We have not paid any attention to the specific questions that arise when we consider the design of political institutions, or the reasons and obligations of political actors. Yet utilitarianism aspires to apply to the political realm too. Indeed, Bentham and his followers in the nineteenth century were arguably more concerned with political matters than with matters of private morality.

⁴⁹ Temkin 2012: 518 and *passim* suggests that no coherent theory may be possible.

On one conception, political philosophy is just the application of general moral principles to the specific issues that arise in politics. If that were true, one might expect utilitarianism to be roughly as plausible as a political philosophy as it is as a moral philosophy. But in fact it is quite common to think that it is more plausible in one sphere than the other.

According to one current of thought, utilitarianism is more plausible as a political philosophy than as a theory of morality. Robert Goodin claims that the features of utilitarianism that many find objectionable when applied to private moral decisions emerge as virtues when applied to political decisions. For example, political decisions should be taken in a calculated and impartial way. Furthermore, limitations of information and the bluntness of public policy instruments mean that good political decisions focus on standard cases, which tends to make utilitarian verdicts more plausible. In sum, in the 'special conditions that characterize public policy-making, utilitarianism looks distinctly credible, in a way it might not for private individuals in guiding their personal conduct'.⁵⁰

This asymmetry makes sense if we think that political issues are distinctive in some important way. There is a long tradition of political thought according to which political decisions operate within different constraints than personal moral decisions, and so are properly governed by different principles. We find this idea, for example, in Machiavelli's contrast between virtue in ordinary life and virtue in political life, and in Max Weber's discussion of the difference between private and political ethics. According to Weber, the role of violence in politics explains the difference.⁵¹

In much contemporary political philosophy we find a similar emphasis on the distinctiveness of politics, but combined with the opposite judgement about the plausibility of utilitarianism. Contemporary political philosophers, echoing Weber, emphasize the fact that politics involves making decisions that are coercively enforced. Furthermore, political decisions are coercively enforced on a population consisting of people who disagree with each other,

⁵⁰ Goodin 1995: 77. I am grateful to an anonymous reviewer for Oxford University Press for encouraging me to consider this current of thought. Bernard Williams reaches the opposite judgement: utilitarianism's 'simple-mindedness' might be all right in private life, '[b]ut the demands of political reality and the complexities of political thought are obstinately what they are, and in face of them the simple-mindedness of utilitarianism disqualifies it totally' 1973: 150.

⁵¹ 'Can the fact that politics operates with a quite specific means, namely power, backed up by the use of *violence*, really be a matter of such indifference as far as the ethical demands placed on politics are concerned?' Weber 1994b/1919: 357, emphasis in the original. For critical discussion of the common interpretation of *The Prince* as asserting the autonomy of politics from ethics see Viroli 2014.

in reasonable ways, about ethical matters. These two features of political decisions seem to many to give rise to a distinctive normative requirement. According to this sort of view, political institutions and decisions must meet normative standards that are distinct from the standards of rightness offered by moral theories.⁵²

One way of marking the distinctiveness of politics is to point to the special importance in political contexts of legitimacy. We routinely distinguish between the rightness of a political decision and the legitimacy of the institution or process from which it issues. Illegitimate processes can issue in right decisions, and vice versa. Note that this enables people to agree about the legitimacy of a decision when they disagree deeply about its rightness. In this way, the distinctness of legitimacy from rightness is essential for the concept of legitimacy to perform its function of providing a way of responding to disagreement that does not depend on resolving the underlying substantive issues.

Thus if utilitarianism were unable to offer a plausible theory of political legitimacy, that would be a further problem with it. Utilitarianism would then be at odds with our convictions about political matters. As it stands, most presentations of utilitarianism do not include discussion of legitimacy. They present it as a theory of the rightness of actions, and perhaps as a theory of the value of laws and institutions. The distinctive concern with legitimacy that is prominent in contemporary political philosophy seems absent from most defences of utilitarianism.⁵³ This may help to explain why utilitarianism appears to have an even lower standing among political philosophers currently than it does among moral philosophers.

Worse, the prospects for developing a utilitarian theory of political legitimacy seem dim. Legitimacy seems to be a procedural property, and as such to be at odds with utilitarianism's focus on outcomes. One instance of this mismatch rises to the surface in utilitarian discussions of democracy. Democracy too is a procedural matter. There is of course a long tradition within utilitarianism of pointing to the benefits of democratic procedures. These include, for example, providing incentives to legislators to rule in the

⁵² For example, Rawls distinguishes between his 'political' conception of justice and 'comprehensive' moral doctrines such as utilitarianism. He writes: 'the principle of utility... is usually said to hold for all kinds of subjects ranging from the conduct of individuals and personal relations to the organization of society as a whole... By contrast, a political conception tries to elaborate a reasonable conception for the basic structure [society's main institutions] alone and involves, so far as possible, no wider commitment to any other doctrine' 1996: 13.

⁵³ One exception is Binmore 1998, though he understands legitimacy differently than most political philosophers.

public interest, gathering information about preferences, and educating voters.⁵⁴ But these instrumental defences of democratic procedures may seem too contingent, and not up to the task of explaining the importance in principle of taking decisions democratically, or of sticking with them when they diminish utility. The technocratic tendencies of utilitarianism seem apt to resurface. More generally, utilitarianism may seem poorly placed to account for procedural normative requirements, such as that political decisions be taken legitimately.

There are other worries about utilitarian approaches to political philosophy. One concerns justice in the allocation of resources. This is closely related to the discussion of the separateness of persons in Section 2.4 above. On the face of it, utilitarianism recommends highly unequal and intuitively unjust allocations of resources in some circumstances. Another concerns rights and liberties, which seem particularly important in the political sphere but which utilitarianism may seem ill-equipped to vindicate. A third concerns the requirement of publicity. Rawls claimed that a satisfactory conception of justice must consist of principles which are such that 'general awareness of their universal acceptance should have desirable effects and support the stability of social cooperation'.⁵⁵ Utilitarianism not only does not subscribe to such a requirement as a matter of principle, but seems impelled to flout it when doing so has good consequences, as Sidgwick noted.⁵⁶

A satisfactory utilitarian political philosophy will have to respond to these challenges. In particular, it will have to offer a satisfactory account of legitimacy and of justice, including explaining whatever is compelling about democracy and about the requirement of publicity. In fact, the idea that distinctively political requirements such as legitimacy and democracy are susceptible to an instrumental explanation is attractive on the face of it, since it is quite plausible to think of political institutions as being instruments for organizing our collective lives. As I shall argue in Chapter 8, the defect has been in not considering sufficiently nuanced and complex forms of utilitarianism in these areas.

⁵⁴ James Mill 1788/1820 emphasizes the importance of frequent elections in bringing the incentives of legislators more closely into line with those of the general public. John Stuart Mill 1791c/1861: 254 emphasizes the educative role of democracy. Riley 1990 claims that elections may play a vital epistemic role by eliciting and aggregating voters' preferences.

⁵⁵ Rawls 1999a: 115.

⁵⁶ Sidgwick 1907: 490. This feature of utilitarianism is criticized in Williams 1973: 135–50.

2.6 Psychology

Utilitarianism has a reputation for psychological flatness and crudity: for failure to understand the complexity of human motivation in its Gradgrindian obsession with measurable facts.⁵⁷ The association of utilitarianism with an impoverished understanding of human psychology is, no doubt, due partly to Bentham's unconvincing attempts to catalogue human motivations and to demonstrate that they reduce to the desire for pleasure and to avoid pain. It was given impetus by John Stuart Mill's anonymous—though ultimately famous—claim that Bentham had only the most superficial and vulgar understanding of psychology.⁵⁸ The question is whether this association is merely an accident of history, or instead reflects an essential characteristic of utilitarianism.

It may at first seem that this can hardly be a serious challenge. As we have defined it, utilitarianism is a family of views united by the attempt to explain ethical phenomena in terms of well-being. It is thus in the business of making claims about evaluative and normative matters, and not in the business of making empirical claims about psychological matters. How could it fail to be compatible with whatever the best psychological theory turns out to be?

A satisfactory moral and political theory should cohere with what we know about human psychology as it is—including ways in which motivations might change—and with our convictions about the inner life of virtuous agents. A theory could fail to do this if, in order to live up to its prescriptions, agents would have to possess superhuman powers, or would have to be motivated in ways that we think incompatible with virtuous or otherwise valuable motivation. Utilitarianism can seem to fail both of these tests. It may seem to require agents to make calculations that are beyond

⁵⁷ 'With a rule and a pair of scales, and the multiplication table always in his pocket, sir, ready to weigh and measure any parcel of human nature, and tell you exactly what it comes to. It is a mere question of figures, a case of simple arithmetic' Dickens 2003/1854: 10. F. R. Leavis comments that 'the justice of this vision of the tendency of James Mill's kind of Utilitarianism, as manifested in later history, can hardly be questioned' 1950: 35.

⁵⁸ 'He saw . . . in man little but what the vulgarest eye can see; recognised no diversities of character but such as he who runs may read. Knowing so little of human feelings, he knew still less of the influences by which those feelings are formed: all the more subtle workings both of the mind upon itself, and of external things upon the mind, escaped him; and no one, probably, who, in a highly instructed age, ever attempted to give a rule to all human conduct, set out with a more limited conception either of the agencies by which human conduct *is*, or of those by which it *should* be, influenced' Mill 1950: 63, emphasis in the original. Mill was, perhaps, being unfair to Bentham. For a more sympathetic portrait, see Schultz 2017: ch. 2.

human powers, and to do so in a way that is incompatible with virtuous concerns and relationships with others and with the phenomenology of ethical life.

The starting point for this thought is that utilitarianism requires agents to make a utility calculation whenever they make a decision. Call this *pervasive calculation*. If utilitarianism requires pervasive calculation it crowds out many valuable features of mental life, including spontaneity, sincere emotional response, and being carefree. Worse, it seems in principle to paralyse decision-making. If every decision must succeed a utility calculation, and allocating time to calculation is itself a decision, then every decision would require an infinite number of prior decisions. No decision could be made.

These bad consequences of pervasive calculation follow no matter what the values employed in the calculation. Once we add that the calculation should be about what maximizes overall well-being, things (if possible) get even worse. For then the activity of calculating seems incompatible with specific partial patterns of concern that are essential to such things as friendship and integrity. It is hard, perhaps impossible, to be friends with someone who always decides what to do based on calculating what would maximize the impartial good.⁵⁹ Moreover, pervasive impartial calculation would alienate the agent not just from others but also from herself.⁶⁰ Finally, if utilitarianism required pervasive calculation it would be hard-pressed to account for the difficulty of ethical decision-making: though calculations can be difficult, they are not difficult in the specific way that generates the anguish typical of conscientious decision-making.

Of course, utilitarianism does not require pervasive calculation. Since calculation is itself an action, utilitarianism requires agents to calculate rightly. We have just noted some very bad consequences of pervasive calculation. So utilitarianism cannot require it.

Calculation is just one way of coming to act. In the standard language, it is just one possible *decision procedure*. The mistake of assuming that utilitarianism requires pervasive calculation is the result of failing to distinguish between a decision procedure and what is known as a *criterion of rightness*.⁶¹

⁵⁹ For discussion of whether consequentialists can be genuine friends see Railton 1984, Coking and Oakley 1995, and Mason 1998.

⁶⁰ See Williams 1973: 93–118 and Railton 1984. For critical discussion of Williams's claims about utilitarianism and integrity see Conly 1983 and Ashford 2000.

⁶¹ Austin 1832: 113–20 made this distinction, and illustrated it memorably: 'It was never contended or conceited by a sound, orthodox utilitarian, that the lover should kiss his mistress with an eye to the common weal' (118). Sidgwick 1907: 413 also draws the distinction very clearly.

These names are potentially misleading. A ‘procedure’ sounds like a recipe for decision-making: a set of rules to be consciously consulted and followed in deliberation; but the intended idea is often much broader than that, and includes for example spontaneous or intuitive decision-making. On the other hand, a ‘criterion’ sounds like something to be used in measurement or assessment. In fact, the intended idea is not connected to anything resembling measurement. A ‘criterion of rightness’ is a specification of what makes actions right or wrong, and not in the first place a specification of how to tell which actions are right or wrong—though, of course, it may have implications for that. Still, we will stick to the established terms for these ideas.

So we should be careful not to assume that because Act Utilitarianism, for example, claims that what makes an action right is that it maximizes utility, it follows that Act Utilitarianism instructs agents to calculate which of their options maximizes utility, and then to perform that option. That decision procedure would, as we noted above, in general fail to maximize utility.

However, this point is not sufficient to undermine the objection from psychology. For one thing, we want an ethical theory to give us some guidance about how to take decisions. That is not the only purpose of an ethical theory; theoretical understanding and retrospective assessment are also important. But it would be disappointing if an ethical theory left us completely in the dark about how to confront the choices we face. For another thing, though, ethical theories inescapably have implications about decision-making, since the activity of taking a decision is sometimes itself an action. So even a theory that downplays the importance of practical guidance will be committed to some claims about the rightness and wrongness of different ways of taking decisions.

For both reasons, utilitarians must go further than simply distinguishing criteria of rightness from decision procedures. They should also try to say something about which ways of coming to act—of taking decisions, in a broad sense that need not involve conscious deliberation—their theory recommends. This will involve trying to specify decision procedures that seem likely to cause the agent to act rightly by their theory’s lights. Figuring out which decision procedure is best requires identifying alternative decision procedures, figuring out the consequences of each, and evaluating them in light of those consequences. These are extremely demanding tasks, and at the very best we can hope to give only rough answers.

Utilitarians also have to hope that the resulting decision procedures cohere tolerably with our sense of how virtuous agents would be motivated.

This is a matter, in the first place, of hoping that the best decision procedure will include things like being motivated by concern for one's friends, by a sense of justice, and the other things we take to be virtues. In addition, it is a matter of heading-off the suspicion that these patterns of motivation are somehow self-effacing or incoherent with the underlying utilitarian theory.⁶²

2.7 Conclusion

These six objections set the agenda for the rest of this book. We are looking for a version of utilitarianism that offers plausible responses to the objection from abhorrent actions, to the objection from demandingness, to the objections about value, and to the objection from the separateness of persons, and which includes a plausible account of political reasons and obligations and a plausible psychology.

I will argue that there is indeed a form of utilitarianism with plausible responses to these challenges. Of course, the plausibility of a reply is a matter of judgement, and such judgements should be appropriately tentative and subject to revision. We should remember both that there are other objections to utilitarianism, and that the rivals to utilitarianism face objections of their own. Still, my aim is to show that utilitarianism remains a worthwhile programme in ethical theory. To do that, I will elaborate in detail the responses that I find most plausible, while also indicating some alternative ways of developing utilitarian theories.

⁶² See Walzer 1973: 171–4. One version of this suspicion is that indirect forms of consequentialism are in some way incoherent. See, for example, the discussion of the incoherence objection to Rule Consequentialism in Hooker 2000: ch. 4 and in Chapter 5, Section 5.4, this volume.

3

Basic Ideas

The version of utilitarianism developed in this book emphasizes the importance of reasons for action. Some presentations of utilitarianism scarcely mention reasons at all, and instead make claims only about the connections between the rightness of actions and the goodness of acts, rules, or motives. But much of our ethical lives is spent thinking about reasons; it is not as though we care only about rightness and goodness. So a satisfying form of utilitarianism should tell us what reasons we have, or else it will fail to account for this important part of what we care about.

The version of utilitarianism I will offer not only ventures to say what reasons we have, but gives them a central explanatory role. Rather than explain rightness directly in terms of goodness, it attempts to explain rightness in terms of reasons, and reasons in terms of goodness. As I shall try to show, this approach has important theoretical advantages. Nevertheless, it amounts to a specific way of configuring basic ethical concepts, and this configuration rests on some controversial assumptions. There are other ways of conceiving the relationships between rightness, reasons, and goodness.

This chapter explains these basic assumptions and offers some motivation for them. It also introduces some of the main ways in which, if we accept these basic assumptions, different versions of utilitarianism can be developed. The purpose is to highlight the wide range of theoretical options consistent with a utilitarian approach to ethics, and to explain how the theory developed in this book is situated within this range of options.

3.1 Reasons

We are interested in a specific kind of reason, namely normative reasons for or against actions. These are to be contrasted on one hand with other normative reasons, and on the other hand with other reasons for action.

Normative reasons are considerations in favour of, or against, some response. They tend to make the response appropriate or inappropriate.

Normative reasons seem to apply to many different kinds of response. For example, the fact that Bernadette is courageous makes it appropriate to admire her. Her courage is a normative reason in favour of the ‘response’ of admiring her. On the other hand, her forgetfulness makes it inappropriate to trust her recollection of events. Here, the fact that she is forgetful is a normative reason against the ‘response’ of trusting her recollection.

Actions are a kind of ‘response’, and as with other responses there can be normative reasons in favour or against them. Suppose that I need to get the antidote out of the safe, and I know that the code is 3652. Then I seem to have a normative reason in favour of entering this code, and a normative reason against entering any other code. It is appropriate for me to enter that code, given the need to get the antidote and the fact that this is the code.

Not all reasons for action are normative reasons. We also have the concept of motivating reasons for action—something that moves an agent to act, because it seems to her to speak in favour of the action. To cite a motivating reason is to offer a particular kind of explanation of action: roughly, it is to point to something that, it is said, caused the agent to act through the mechanism of making it seem to her a good thing to do. Thus motivating reasons are, necessarily, *perspectival*: something cannot be a motivating reason for an agent unless it makes sense of the action from the agent’s perspective.¹

One of the main controversies about normative reasons for action is whether they too are perspectival in this sense, or some related sense. But whether they are or not, the following is clear. While it seems possible for agents to be motivated by considerations genuinely in favour of (or against) actions—that is, by normative reasons for or against actions—it also seems possible for agents to be motivated by considerations that are not genuinely in favour of (or against) actions. Suppose I buy a lottery ticket because I feel lucky and I believe that this makes me more likely to win. Though feeling lucky motivated me, it is not a consideration in favour of buying a ticket, since feeling lucky does not make me more likely to win.² Even if normative reasons are perspectival in some way, not all motivating reasons for action are also normative reasons for action.

¹ For a recent discussion of some different views about the nature of motivating reasons see Alvarez 2018. She characterizes the concept of motivating reasons like this: ‘When we are interested in what reasons motivate an agent to act, we are interested in the reasons that the agent takes to favour her action, and in the light of which she acts’ (3298).

² More precisely, it is not a reason for buying a ticket that has to do with the chance of winning. It could be a reason of a different kind. For example, feeling lucky might involve having a desire to buy a ticket. There might be a reason to satisfy this desire that has nothing to do with the chance of winning.

In what follows, our interest will be in normative reasons for or against action, and I shall refer to them as ‘reasons’, for short.³ An important feature of reasons is that they are *contributory*.⁴ A reason in favour of an action, as we said before, ‘tends’ to make that action appropriate. It need not make it flat out appropriate, or right. It is a commonplace but very important feature of life that it is possible for there to be a reason in favour of an action and yet stronger reasons against it, or vice versa. For example, the fact that the prank would be fun is a reason in favour of it, but its cruelty may well be a stronger reason against. This means that, in general, it is illicit to conclude that an action is appropriate or right just because there is a reason in favour of it.

We want to know not just what reasons there are, but also how they interact with each other. One element of the interaction of reasons has to do with their strength or ‘weight’.⁵ We are used to thinking of some reasons as stronger than others. For example, there may be a strong reason to tell an hilarious joke on some occasion. Doing so would put everyone at ease and cause much merriment. On the same occasion there may be some reason, though a much weaker one, to tell a feeble joke. After all, it would bring some small amount of pleasure.

It is likely that many or even most of our reasons are very weak. Like the reason to tell a feeble joke, they are weak considerations for or against actions. Since reasons are only contributory, agents should often ignore the very weak ones.⁶ Only rarely will they make any difference to what we ought to do.

The utilitarian approach to explaining reasons is *teleological*, since it seeks to explain reasons in terms of goodness.⁷ Act Utilitarians, for example, seek

³ Reasons for other kinds of response—such as reasons for attitudes or feelings—are also important in ethics. However, with the exception of the discussion of virtue in Chapter 9 they will not be the main focus in this book.

⁴ Dancy 2004: 15–16.

⁵ Reasons might interact in ways more complex than can be captured by the metaphor of summing their weights. Jonathan Dancy claimed that ‘reasons are like rats, at least to the extent that two rats that are supposedly on the same side may in fact turn and fight among themselves’ 2004: 15. For the time being we will leave open the issue of how reasons interact, returning to it in Chapter 9.

⁶ Schroeder 2007a: 92–97 argues that in general we should be cautious about ‘negative reasons existentials’, i.e. claims that reasons do not exist, because we presuppose in discussions of reasons that only relatively strong reasons will be mentioned. This presupposition tends to make us overlook weak reasons, and to judge that, where only weak reasons exist, there are no reasons at all. He claims that each of us has, for example, an extremely weak but genuine normative reason to eat a car, due to the fact that ‘it contains the recommended daily allowance of iron’ (96).

⁷ Portmore 2011: 59–62 claims that the teleological conception of reasons should be formulated in terms of ‘reasons to desire’ outcomes, not in terms of the ‘goodness’ or ‘desirability’ of

to explain reasons in terms of the goodness of acts themselves. According to them, the fact that an action would be good is a reason to perform it. Rule Utilitarians offer a different explanation, in terms of the goodness of rules, and Motive Utilitarians offer a different explanation again, in terms of the goodness of motives. But these are still teleological theories of reasons, since they seek to explain reasons in terms of the goodness of something or other.

The teleological approach to explaining reasons suggests one way of accounting for the strength of reasons. Since strength and goodness both come in degrees, it seems natural to try to explain the strength of reasons in terms of the degree of goodness with which they are associated. This is a natural way to explain why the reason to tell the hilarious joke is stronger than the reason to tell the feeble joke: it leads to a better outcome. Thus teleologists are quite likely to claim that strength of reasons reflects degrees of goodness.⁸ They may go further, and claim that strength is the only factor on which the interaction of reasons depends. However, it is important to note that this second claim is optional for teleologists. It is possible to think that reasons are ultimately explained in terms of goodness, whilst also thinking that reasons interact in more complex ways. This could happen if there are reasons associated with moral rights, for example.⁹

In common with other forms of teleology, utilitarian theories of reasons can be contrasted with some other important approaches to ethics. According to 'reasons-first' approaches, we should instead try to explain goodness in terms of reasons. For example, we might say that for something to be valuable is for it to have properties that provide reasons for valuing it or for other positive responses towards it.¹⁰ Alternatively, according to 'fittingness-first' approaches, we should try to explain reasons and goodness in terms of fittingness. We might then say that for something to be valuable is for it to be fitting to value it, and for something to be a reason for a response is for it to

outcomes. Portmore's worry is that the concepts of goodness and desirability exclude agent-relativity. As I will make clear shortly, however, the teleological theory is best understood as tying reasons to distinct goods, which can include things that are good for someone but neutral or bad for everyone else. This introduces a kind of subject-relativity of reasons. Moreover, as I shall explain in Chapter 5, agent-neutral goodness can generate agent-relative reasons.

⁸ For present purposes we can ignore the relevance of subjective or objective probability to strength of reasons. Assume that whatever significance these kinds of probability have is included in our measure of the goodness of actions.

⁹ We will return briefly to this issue in Chapter 9.

¹⁰ For example, see Scanlon 1998: ch. 2.

explain why that response is fitting.¹¹ These other approaches each describe a very broad family of views with which standard utilitarian views are in competition.¹²

Second, all utilitarian theories of reasons reject standard forms of ‘internalism’ about reasons. According to standard forms of internalism, something can be a normative reason for an agent only if she could, via some appropriate process, be motivated to act on it.¹³ That is, these views claim that there is an ‘internal’ connection between motivating and normative reasons, such that something cannot be a normative reason for an agent unless it could motivate her. In contrast, utilitarians typically claim that everyone has a normative reason to promote others’ well-being whenever they can do so. On the assumption that some people could not be motivated by the fact that something would benefit another, utilitarian theories are incompatible with these forms of internalism.¹⁴ Consequently, they are incompatible with every theory of reasons that accepts these forms of internalism, including Humean theories of reasons.¹⁵

One source of opposition to the utilitarian approach to explaining reasons is the claim that it appeals to a concept of goodness that we cannot understand. According to this line of criticism, utilitarianism appeals to the concept of goodness *simpliciter*—that is, the concept of something’s being, simply, good, rather than good in a way or good for someone. Like others, Judith Jarvis Thomson doubts that we understand this concept. She writes:

[I]f we do not know in what way a man means that a thing is good when he says of it ‘That’s good’, then we simply do not know what he is saying of it. Perhaps he means that it is good to eat, or that it is good for use in making cheesecake, or that it is good for Alfred. If he tells us, ‘No, no, I meant that it is just plain a *good thing*,’ then we can at best suppose he is a philosopher making a joke.¹⁶

¹¹ For example, see Ewing 1939, Chappell 2012, McHugh and Way 2016, Howard forthcoming.

¹² It is possible to combine utilitarianism with a reasons-first or fittingness-first approach. One way to do this would be to explain rightness of actions and reasons *for action* in terms of the goodness of states of affairs, but to explain the goodness of states of affairs in terms of reasons *to desire*, or the fittingness *of desiring*, those states of affairs. See Chappell 2012 for relevant discussion.

¹³ Williams 1981b. For a criticism of Williams’s way of framing the issue, see Hurley 2001.

¹⁴ There are different forms of internalism, specifying the ‘appropriate process’ differently.

¹⁵ Smith 1994 and Schroeder 2007a are two prominent recent discussions of Humean theories.

¹⁶ Thomson 1997: 276, emphasis in original.

As Thomson points out, this threatens consequentialist theories (including utilitarianism) with incoherence.¹⁷ This is because consequentialist theories seem to rely on the concept of overall goodness of outcomes, and overall goodness appears to be a kind of goodness *simpliciter*.

However, this line of criticism is misplaced. For one thing, it is arguable that we do already have a concept of goodness *simpliciter*. Conor McHugh and Jonathan Way offer one way of characterizing it, when they note that we could take it to be what we have reason to value for its own sake.¹⁸ Second, even if we do not already have a concept of goodness *simpliciter*, the utilitarian could stipulatively introduce one to suit her purposes. For example, she might define goodness *simpliciter* as the sum total of well-being.

Much more importantly, though, utilitarians do not need to appeal to goodness overall (and so, goodness *simpliciter*) in their theories of reasons.¹⁹ This may seem surprising. But it is much more plausible—and it is consistent with utilitarianism—to say that the source of reasons is the well-being of each subject, than to say that it is goodness overall. If I face a choice between benefiting Jane slightly or Jack greatly, I have some reason to benefit Jane, associated with her well-being as a distinct good, even though I have a stronger reason to benefit Jack, associated with his well-being as a distinct good. We should not say that I have only a reason to bring about the outcome that is best overall, i.e. to benefit Jack.²⁰ In fact, we should go further than this, and associate utilitarian reasons with distinct parts of each subject's overall well-being. If I can either benefit Jack slightly now or benefit him greatly later, I have some reason to benefit him now even though I can benefit him more by not doing so.

This strongly suggests that utilitarians should claim that reasons flow from the distinct benefits and harms that can be brought about, and not (to avoid double counting) from the overall goodness of outcomes. This means that they should explain reasons in terms of subjects' well-being, which is

¹⁷ Thomson 1997: 276. As she acknowledges, Thomson is here pursuing a line of thought developed in Geach 1956 and Foot 1985. Geach claimed that 'good' is an 'attributive' rather than a 'predicative' adjective, in the sense that 'there is no such thing as being just good or bad, there is only being a good or bad so-and-so' (34). Byrne 2016 is a recent discussion of the idea of predicative goodness.

¹⁸ McHugh and Way 2016: 578. They cite this as an attractive feature of the buck-passing account of value. However, one need not accept the buck-passing account of value to accept a conceptual connection between reasons and value, which is what is being used here to characterize goodness *simpliciter*.

¹⁹ They may have to appeal to goodness *simpliciter* in other parts of their theories, such as in their theories of virtue.

²⁰ See Chapter 2, Section 2.4 above and Chappell 2015.

goodness-for, rather than in terms of goodness overall. Reasons are not for goodness' sake. The concept of goodness overall—and so of goodness *simpliciter*—plays no essential role in the best utilitarian theories of reasons.

As we have already noted, a very important issue for all theories of normative reasons is whether they are perspectival in some way. For utilitarians, this issue is the question of whether the relevant concept of goodness in their theories is *actual* or *expected* goodness. Here, 'expected goodness' means 'goodness expected from some relevant perspective'. The relevant perspective is not likely to be defined as, in all cases, the agent's actual perspective, since we do not want to say that feeling lucky gives me a reason to buy a lottery ticket, even though I expect to win. Some degree of idealization from the agent's actual perspective seems necessary. On the assumption that the idealization does not extend to assuming omniscience, however, actual and expected goodness will diverge in most cases.²¹

It makes a great difference to a theory of reasons whether it characterizes reasons as perspectival. This is especially so for consequentialists, for whom the 'consequences' of an action include everything that would happen were the action performed, including events in the distant future. Later, I shall claim that normative reasons are not perspectival. Before we get to that, though, we should consider some of the other basic concepts of ethics and their relationships to reasons.

3.2 Rightness

Like the concept of reasons, the concept of the rightness of actions appears to be central to our ethical thought.²² When deliberating, we often seem to be trying to figure out which of the actions we could perform would be right, and which would be wrong. When retrospectively assessing actions—our own or others'—we often use the concept of rightness. When considering questions of praiseworthiness and blameworthiness, many think that the question of whether an action was right or wrong is very important. Moreover, these issues pack an emotional punch. When someone concludes that he has acted wrongly, he is likely to feel some strong emotion, such as

²¹ I leave open exactly how the relevant perspective should be specified; for present purposes that does not matter. For discussion, see Jackson 1991: 463–5.

²² Anscombe 1958: 1–2 claims that the concept of *moral* rightness is not central to Aristotle's thought. For doubts about this, see Crisp 2006: 31–5.

shame, guilt, or remorse. In these and other ways, the concept of rightness seems to loom large.

It does not follow that the best ethical theory will find room for it. Though it appears central to much of our ethical thought and practice, we might be better off without it. We shall consider some claims to that effect shortly. If, as I shall claim, we should retain the concept of rightness, then difficult questions remain about how to understand its relationship with the concepts to which it seems most closely connected: the concept of reasons, the concept of deliberation, and the concepts of praiseworthiness and blameworthiness. These issues will occupy us in this section and the following two sections of this chapter.

The concept of rightness is, in some respects, quite clear. To say that an action is right is to say that it is not wrong, and there seem to be at least two different ways in which an action may be not wrong. First, it may be ‘required’, which means that not performing the action would be wrong. Second, it may be ‘optional’, which means that performing the action would not be wrong and that not performing it would also not be wrong. Summarizing, we can say that an action’s being right or wrong consists in its having one or other of the three kinds of ‘deontic status’ shown in Table 3.1.²³

Admittedly, this does not get us terribly far. It usefully disambiguates ‘right’, since that is sometimes used to mean ‘required’ and sometimes used to mean ‘required or optional’ (i.e. not wrong). It also does this in a way that is suitably general. This way of understanding talk of rightness applies not only to talk of moral rightness but also to talk of legal, prudential, comedic, architectural, or gastronomic rightness. Still, its explication of the three kinds of deontic status has us travelling around a fairly tight circle, since

Table 3.1. Three kinds of deontic status of an action

| Deontic status of action X | Explication in terms of wrongness |
|----------------------------|--|
| X is required | Not to do X would be wrong |
| X is optional | To do X would not be wrong, and not to do X would not be wrong |
| X is prohibited | To do X would be wrong |

²³ Some draw further distinctions within the category of the optional: *supererogatory* actions are not required but are better than what is required; *suberogatory* actions are not prohibited but are still morally deficient. On the supererogatory, see Vessel 2010. On the suberogatory, see Driver 1992: 291. Driver cites failing to donate a kidney to one’s sick brother as an example of a suberogatory act (287).

each explication is in terms of wrongness, which is very closely related to the concept of rightness.

We can get a little further by noting two other features of the concept of rightness. One is that, like ‘goodness overall’, it makes an *overall* assessment. In this respect it is unlike the concept of a reason, which as we noted earlier is a contributory concept. To say that an action is prohibited, or optional, or required, is to offer some kind of overall assessment of it, not to point to just one of its qualities or defects. This does not mean that, when we say that some action is right, we mean to take into account every one of its qualities. We may have some restricted range of qualities in mind, as when we say that ‘comedically that is the right thing to do’. But, with respect to that range, we mean to assert some overall assessment of the action.

Second, unlike ‘goodness overall’, rightness is an *all-or-nothing* rather than a *scalar* concept. We do not usually think that there are *degrees of prohibitedness*, even though we do think that some prohibited actions are worse than others. The same goes for an action’s being required or optional: these kinds of deontic status appear to be all-or-nothing matters, in the sense that they do not admit of degrees.²⁴

It is clear that we need a way of assessing actions overall in our ethical repertoire. In deciding what to do, for example, it would be helpful but not sufficient to know that there is a reason in favour of a certain option. Since there may be a decisive reason against it, knowing that there is a reason in favour would not licence any conclusion about whether to perform it. In order to reach a conclusion of that sort we need some way to assess the overall picture.

The question, then, is whether we should want an all-or-nothing overall assessment of actions, of the sort that the concept of rightness involves. According to Scalar Utilitarians, we should not. They claim that the best form of utilitarianism limits its overall assessment of actions to assessment of their overall goodness. Thus, it may claim that one action is better overall than another, but it makes no all-or-nothing claims such as that one action is required, or another is prohibited, or another is optional.²⁵

One problem with Scalar Utilitarianism is simply that it is highly revisionary. Many of our ethical convictions are convictions about rightness. For

²⁴ Mill 1998/1861: 55 (ch. 2) uses the phrase ‘right in proportion’. However, this strikes most as deviant usage, and Crisp 2014: 232 doubts that Mill means it literally. Peterson 2013 treats rightness as a matter of degree; Crisp 2016: 39–44 criticizes Peterson.

²⁵ See Norcross 2006 and Tobia 2017. Though Scalar Utilitarianism is usually presented as a version of Act Utilitarianism, it need not take that form. A Scalar Rule Utilitarian might try to rank actions according to the goodness of the sets of rules that permit them, for example.

example, suppose that I could either save two lives by pressing a button, at no cost to myself, or I could do nothing. Most of us have the conviction that I have a very strong reason to press the button, and most of us have the conviction that pressing it is much better than doing nothing. Scalar Utilitarians seem able to account for both of these convictions about this case. But they cannot account for another very strong conviction that most of us have about this case, to the effect that doing nothing would be wrong. Such a conviction is, according to Scalar Utilitarianism, simply a mistake.²⁶

Since it is possible that the best ethical theory is highly revisionary, this objection to Scalar Utilitarianism is not quite decisive.²⁷ So we should ask whether there is a worthwhile function for an all-or-nothing overall assessment of actions to play in our ethical thought. If there is not, perhaps we should engage in the deep revision of our thought recommended by Scalar Utilitarians.

In fact, however, the problem seems to be just the opposite. There are, if anything, *too many* worthwhile functions for an all-or-nothing overall assessment of actions to play—at least, there may be too many for any single assessment to play. First, one way to think of deliberation is as a process of trying to determine which option is right, or for which there is sufficient reason to act. Since the notion of sufficiency involves a threshold, either of these conceptions of deliberation portrays it as employing an all-or-nothing assessment of actions. Second, we engage in retrospective assessment of actions, including when making judgements of praiseworthiness and blameworthiness. Often on these occasions we want to know whether a past action was right or wrong. This could be because we treat wrongness as a necessary condition of blameworthiness, for example, or because we want to evaluate our decision-making and we treat production of right actions as the relevant standard.²⁸ Finally, in thinking about reasons for action we sometimes want to distinguish between an action's being favoured by the strongest reason

²⁶ This objection is Mulgan's 2001: 131, 143. For other objections to Scalar Utilitarianism, see Lawlor 2009, McElwee 2010a, and Lang 2013.

²⁷ Crisp 2006 argues for a revisionary view. Like Anscombe and like Scalar Utilitarians, he tries to do without the concept of moral obligation or moral rightness. But unlike Scalar Utilitarianism, his view includes an all-or-nothing form of overall assessment: his notion of 'overall reason' (56–7) appears to mean 'strongest reason all things considered', which is an all-or-nothing concept.

²⁸ In Chapter 9 I will claim that good deliberation does not involve trying to discern reasons for action or the rightness or actions, and also that wrongness is not a necessary condition of blameworthiness. However, it remains true that these are salient roles for the concept of rightness. If it is claimed that rightness does not play these roles, we are owed an account of what does play them.

and its being favoured by a strong enough reason. Scalar Utilitarianism can provide a proxy for one of these concepts, insofar as ‘best action’ correlates with ‘action favoured by the strongest reason’. But it does not provide a correlate for the other concept: ‘action favoured by a strong enough reason’. Standardly we mark the distinction by distinguishing between those actions that are required and those that are optional. That is, we use the concept of rightness to mark the distinction.

Thus there are at least three candidate functions for an all-or-nothing assessment of actions in our ethical thought: in deliberation, in retrospective assessment of actions, and to mark an important distinction between ways in which reasons favour actions. The question is not whether we can do without an all-or-nothing assessment of actions, but whether the concept of rightness can play all of these roles for us at once. The sort of assessment of actions involved in deliberation and in judgements of praise and blame is perspectival, since it must find some way to take into account how things seem or seemed to the agent. If we also seek to hold on to a non-perspectival concept of rightness, it may be that we cannot employ a single concept in all three roles. The different uses to which we can put the concept of rightness threaten to fragment it.

3.3 Two Ways to Avoid Fragmentation

Brad Hooker outlines and argues for one way of avoiding fragmentation. If we define rightness in a perspectival way, we can maintain the close intuitive connections between the concepts of rightness, good deliberation, and praise and blame. Discussing the best way to formulate Rule Consequentialism, he writes:

If the concept of wrongness wanders too far away from blameworthiness, most (if not all) of the importance of wrongness seems lost. To retain its importance, wrongness has to stick fairly closely to blameworthiness. Thus, wrongness, like blameworthiness, needs to be explained in terms of rules reasonably expected to be best, not rules that actually would be best.²⁹

Hooker allows that there can be some distance between wrongness and blameworthiness, since someone could do the wrong thing but have a valid excuse, and so not be blameworthy. But he thinks that allowing

²⁹ Hooker 2000: 75.

wrongness and blameworthiness to diverge from each other too much risks undermining the interest of the concept of wrongness. To avoid that, he defines wrongness (and rightness) in terms of the perspectival notion of expected goodness, rather than the non-perspectival notion of actual goodness.³⁰

Though he does not mention reasons in these passages, Hooker's proposal allows for the intuitive connection between reasons and rightness also to be maintained, so long as reasons are understood in a perspectival way. On this view there is a single concept of rightness, and it retains all of the connections we have identified. Figure 3.1 summarizes this proposal.

In this diagram, the unbroken lines represent close conceptual connections. For example, on Hooker's proposal there is a close connection between the concept of wrongness and the concept of blameworthiness. Both are specified in terms of how things do or should seem to the agent; the fact that not every wrong action is blameworthy (since wrong actions can be excused) is consistent with the existence of a close connection between the concepts, in the sense I intend. On this understanding, each of the four concepts is closely connected with the other three, and all are understood in a perspectival way.

The great attraction of this view is that it elegantly avoids fragmentation. Rightness retains its connections with reasons, good deliberation, and

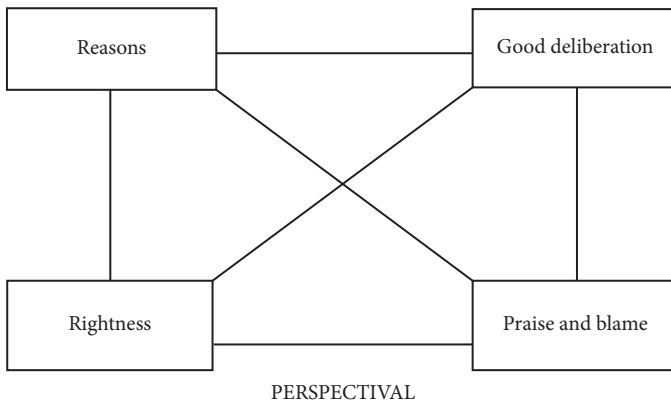


Figure 3.1 Hooker's proposal

³⁰ Jackson 1991: 466–7 argues for a perspectival conception of rightness by emphasizing the connections between right action and good deliberation.

praiseworthiness and blameworthiness. These close connections underline the importance of rightness, as Hooker points out.

However, I shall not adopt Hooker's proposal. The trouble is that at least one concept of reasons and at least one concept of rightness seem to be non-perspectival, and neither is one we should be ready to give up. To see the importance of the non-perspectival concepts of reasons and rightness, consider the following case. Suppose that I am trying to decide whether to move out of the city into the country. I don't know for sure what living in the country is like, or how my life would continue were I to stay in the city—but I make my best estimates of these things, and in doing so I commit no errors of reasoning and ignore none of my evidence. On the best perspectival view of my reasons, let us stipulate, the two options are exactly equal. From the relevant perspective they have equal value, and I have reasons of equal strength to stay or go, and so each option is, from this perspective, right.

Now suppose that, in fact, I would hate living in the country but would get along just fine if I were to stay in the city. These facts are not accessible from the perspective relevant to my reasons and the rightness of my actions, according to the best perspectival view of those things. Yet it is very hard to accept that they do not speak in favour of staying in the city. They seem to be 'considerations in favour' of staying in the city, and against moving to the country. It also seems that, in at least one sense of 'right', these facts mean that staying in the city would be the right thing for me to do, and moving to the country would be the wrong thing. These things seem true even though these facts are not considerations that I can access from my perspective, and so I cannot be criticized or blamed for failing to take them into account.

If this is correct, it supports the claim that there is a concept of 'reason' and of 'right' that is non-perspectival and that we should retain. That claim is further supported by the following variation of the case. Suppose that, remarkably, I come to know these facts about my future as a result of bumping my head. If reasons and rightness are perspectival matters, this change in what I know brings about a *change* in my reasons and in the rightness of my options. Now that I know that I would hate living in the country, I have a strong reason to stay put, and that would be the right option for me. But that does not seem to describe my situation correctly. There is, at least, a sense of 'reason' according to which bumping my head did not change my reasons, but instead caused me to *see them more clearly*. That must be true, or else this change in my epistemic position would not be welcome so far as my understanding of these reasons and the rightness of

my options goes. It would, instead, merely cause a change in my reasons and the rightness of my options, which all along I have perfectly understood.

We should not try to give up the non-perspectival concepts of reasons and rightness. Giving them up would make it very hard to make sense of why we welcome improvements in our knowledge of the consequences of actions.³¹ As a result, we should reject Hooker's proposal for avoiding fragmentation.

G. E. Moore accepted a non-perspectival concept of rightness, and outlined a different way of trying to avoid fragmentation. Moore claimed that rightness should be defined in terms of actual, not expected, consequences. But he then considered an objection relating to uncertainty. Suppose that an agent is deciding what to do, and he has 'every reason to think' that one course of action will be best. Won't we be inclined to say that he ought to perform this action, even if it turns out not to be best? Moore writes:

It may be admitted that we should say, and should be justified in saying, that he absolutely *ought* to choose the course, which he has reason to think will be the best. But we may be justified in saying many things, which we do not know to be true, and which in fact are not so, provided there is a strong probability that they are. And so in this case I do not see why we should not hold, that though we should be justified in saying that he *ought* to choose one course, yet it may not be really true that he ought. What certainly will be true is that he will deserve the strongest moral blame if he does not choose the course in question, even though it may be wrong.³²

Moore's proposal seems to be as follows. We ought to define rightness in a non-perspectival way, as depending on actual consequences. However, we can define good deliberation, praiseworthiness, and blameworthiness in terms of justified beliefs about rightness. We can then say that the agent in Moore's example is deliberating well, and does not deserve blame, when he chooses the option that he expects to be best, since he is choosing the option that he justifiably believes is right; and we can say this even if that option turns out to have worse consequences than another option would have had. Moreover, we can and ought to say that he would have been blameworthy

³¹ Perspectival views can make sense of our reasons to *seek* such improvements, as Jackson notes 1991: 464–5. However, it is at least very difficult for them to make sense of epistemic manna: epistemic improvements that just befall us, as with the bump to my head. For an attempt to show that a perspectival view can make sense of this, see Lord 2015: 42–50. For a different argument that there are non-perspectival reasons see Crisp 2006: 39–40.

³² Moore 2005: 101, emphasis in original. Note that here Moore uses 'should' to express the conditional mood, rather than to mean 'ought'.

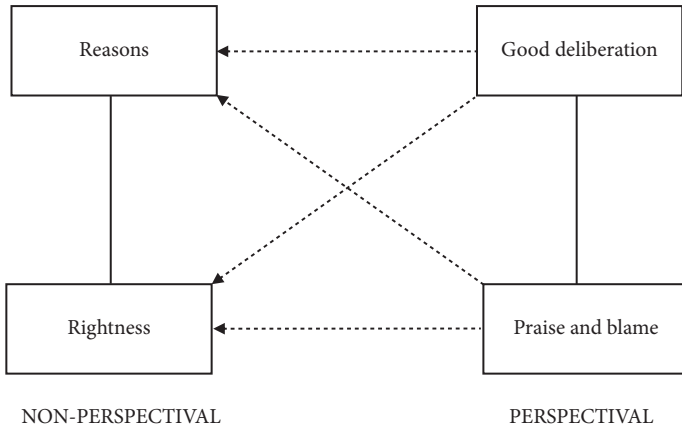


Figure 3.2 Moore's proposal

for choosing differently, even if by luck he were to choose an option with better actual consequences than the one he expected to be best.

Moore's proposal is attractive because it seems to offer a way of avoiding fragmentation while recognizing the non-perspectival concepts of reasons and rightness. Figure 3.2 represents this proposal.

In this diagram, the broken lines represent the somewhat less close connection between (for example) good deliberation and rightness than is a feature of Hooker's proposal. The arrows represent the fact that the items on the right-hand side of the diagram are characterized in terms of *justified beliefs about* the items on the left-hand side.³³

Though Moore's proposal is tempting, it does not work. This is shown by a class of cases that have come to be known as 'Jackson Cases'. In Jackson's example, a doctor must decide between three courses of treatment for a patient with a skin condition. The doctor knows that treatment *A* would provide a partial cure, and she knows that one of *B* and *C* would provide a total cure, while the other would kill the patient; unfortunately, she has no way to find out which of *B* and *C* would cure the patient and

³³ This is to specify the arrows in terms of 'doxastic justification': whether what the agent believes about reasons or rightness is justified. Another way of specifying Moore's proposal would be to have the arrows represent 'propositional justification': whether the agent has good reason to believe the relevant propositions about reasons or rightness. For present purposes we can ignore this distinction, since the reason I give below for rejecting Moore's proposal applies whichever specification we adopt. For discussion of the distinction see Turri 2010.

which would kill him.³⁴ Intuitively, good deliberation would lead the doctor to prescribe *A*, and she would be blameworthy for prescribing *B* or *C*, even if by luck she thereby would prescribe the complete cure. Unfortunately, Moore's proposal conflicts with these judgements. The doctor knows for sure that *A* has worse consequences than one of *B* and *C*. Thus, on Moore's view of rightness she is justified in believing that *A* is not the right option; we can add, if we like, that she in fact believes this. According to Moore, therefore, if she prescribes *A* she 'will deserve the strongest moral blame'. But intuitively she would be blameworthy for prescribing anything other than *A*.

Moore's proposal works for cases of uncertainty in which it is justifiable to believe, of the option with greatest expected value, that it is the option with greatest actual value. In Jackson Cases, however, it is not justifiable to believe this, and if we endorse the standard intuitive judgements about these cases, Moore's proposal fails.³⁵

3.4 Three Ways to Accommodate Fragmentation

Unfortunately, neither Hooker's proposal nor Moore's proposal offers a satisfactory way of avoiding fragmentation between the roles we may expect the concept of rightness to play. We should instead find some way to accommodate this fragmentation.

One natural proposal is to adopt a kind of pluralism about rightness and reasons. According to this kind of pluralist, we should distinguish between perspectival rightness and reasons, on one hand, and non-perspectival rightness and reasons, on the other.³⁶ We can then accommodate the judgements I claimed were intuitive about my choice whether to move to the country: we can say that, in the non-perspectival sense, I have a strong reason (even before the bump to my head) to stay put, and that this would be the right option. We can also explain why good deliberation does not come down in favour of staying, and why if choose to move I am not blameworthy

³⁴ See Jackson 1991: 462–3. Jackson says that the doctor believes that *A* is 'very likely' to be a partial cure, but I am imagining that she is certain about this. Similar cases were discussed in Regan 1980: 264–5 n. 1, and Parfit 1987: 25. For discussion of Jackson Cases and perspectival rightness see Bykvist 2009, Smith 2010b, and Mason 2013.

³⁵ Compare Lord 2015: 31.

³⁶ Smith 2010b: 71 endorses pluralism of this sort, as does Parfit 2011: 151. Jackson is a pluralist of a different kind, recognizing one kind of rightness for each possible perspective 1991: 471–2. Broome 2013: 40 claims that, even if pluralism is true, the 'central ought' is perspectival.

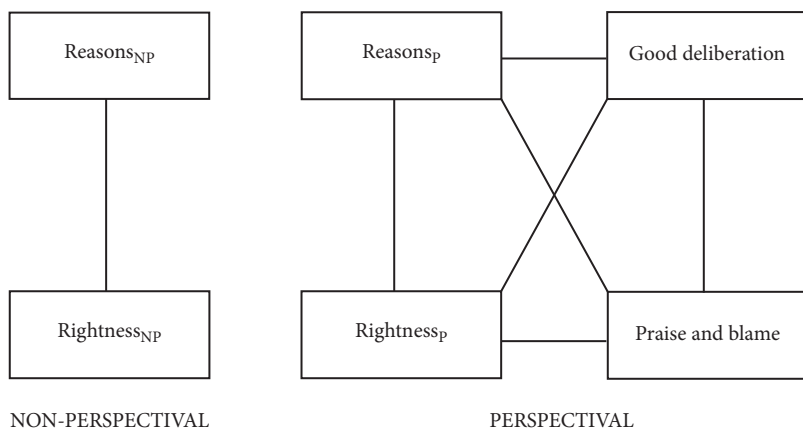


Figure 3.3 Perspectival/non-perspectival pluralism

for this choice, by saying that good deliberation and blameworthiness are concerned with perspectival reasons and rightness. In this way, we assign some functions to the perspectival versions of the concepts, and other functions to the non-perspectival versions. I will call this ‘perspectival/non-perspectival pluralism’. Figure 3.3 represents this view.

On this view, there is no close connection between the perspectival concepts on the right-hand side and the non-perspectival concepts on the left-hand side. As the diagrams indicate, pluralism in effect adopts Hooker’s proposal whilst adding non-perspectival concepts of reasons and rightness.

A second proposal is to dissociate reasons from rightness. We can maintain a close connection between rightness and praiseworthiness and blameworthiness, and between rightness and good deliberation, by denying that the rightness of an action entails anything about the reasons for or against it. On this view reasons are non-perspectival, but rightness is a perspectival matter—as befits its close connection with deliberation and praiseworthiness. To act rightly is to act in a way that is either praiseworthy or at least not blameworthy, and so to act in a way that makes sense from some relevant perspective—such as the agent’s own perspective, or perhaps some idealized version of it. But on this view it does not follow from the fact that some agent acted rightly that she acted on good reasons, or with sufficient reason, or in the way for which there was most reason.³⁷

³⁷ I learned the idea that wrongness might be more closely connected with blameworthiness than with reasons from Brian McElwee, in conversation. Some of his published work makes

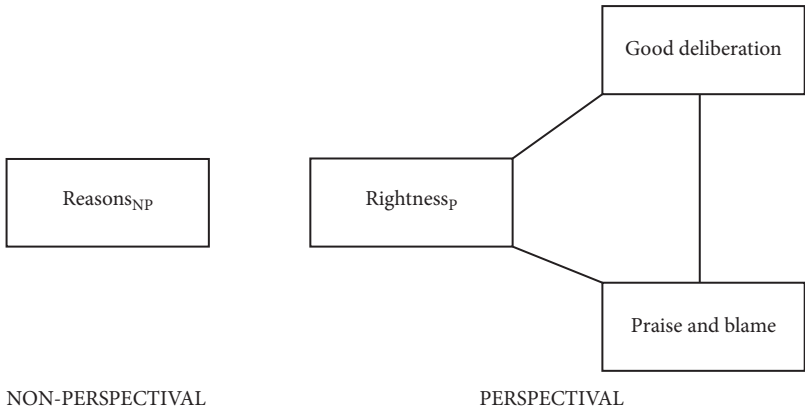


Figure 3.4 Dissociating reasons from rightness

Following this suggestion we could say that I have a strong reason to stay in the city, of which I am not aware. But we could also say that, since wrongness is associated with blameworthiness and not with reasons, I need not be acting wrongly if I choose to move to the country. Blameworthiness, and so rightness, is a perspectival matter, and I was faultlessly unaware of my reason to stay put. This proposal is represented in Figure 3.4.

As this diagram shows, the dissociating proposal is monistic about both rightness and reasons. It allows us to treat rightness as a perspectival matter, and reasons as non-perspectival, by breaking the conceptual connection between them.

A third proposal retains the connection between rightness and reasons, but accepts a simple divorce between these concepts and the concepts of good deliberation, praiseworthiness, and blameworthiness. On this view deliberation and praise and blame are perspectival matters, but they have no close connection with reasons and rightness, which are non-perspectival.³⁸ This view is represented in Figure 3.5.

similar claims: for example, he writes: "To say that one has done something morally *wrong* is not to say that one has done something other than what one had most moral reason to do, but rather to suggest that one would generally merit blame or serious criticism for acting in this way" (McElwee 2010b: 397, emphasis in the original; see also Mill 1998/1861: ch. 5 and Adams 1999: ch. 10. However, the dissociating proposal discussed in the text above is not McElwee's view: he distinguishes different uses of 'right', including a non-perspectival use.

³⁸ My thinking about these issues has been influenced by Bykvist 2009.

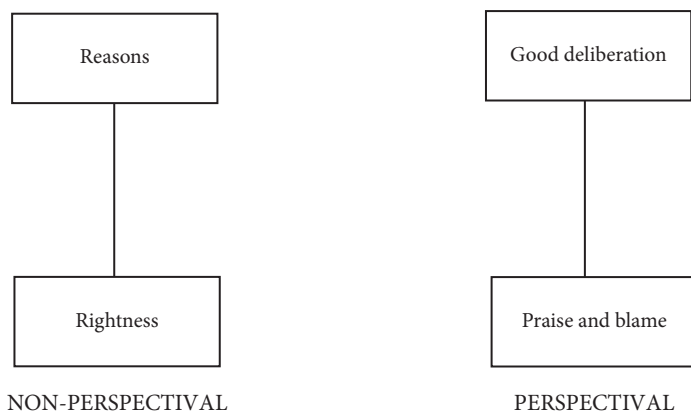


Figure 3.5 The simple picture

In one respect this picture resembles perspectival/non-perspectival pluralism. Neither postulates a connection between non-perspectival reasons and rightness, on one hand, and good deliberation and praise and blame, on the other. But of course it sharply differs from pluralism in not recognizing a further, perspectival, kind of reasons and rightness. This makes it a simpler picture of the basic concepts of ethics. We might worry that, in dissociating rightness from deliberation and praise and blame, it diminishes the interest or importance of rightness, as Hooker feared. We might worry also that it leaves unexplained what makes for good deliberation, and the difference between praiseworthiness and blameworthiness.

Despite these worries, I will adopt the simpler picture. Earlier I argued that we need to retain a non-perspectival concept of reasons and rightness in order to explain why the bump to my head caused me to see my reasons more clearly, rather than causing those reasons to change, and to explain the sense in which, all along, it would have been right for me to stay in the city. That rules out Hooker's proposal, along with the proposal to dissociate reasons from rightness.³⁹ Jackson Cases, so I claimed, rule out Moore's proposal. That leaves pluralism and the simple picture. The choice between these views turns on whether the best explanations of praiseworthiness,

³⁹ Dissociating reasons from rightness also has the disadvantage of leaving us without a ready concept to express judgements about the overall balance of reasons favouring or disfavouring actions. We might assign 'most reason' to this role. However, this concept does not capture the important distinction between actions that are optional and actions that are required. As a result, it fares worse than the simple picture in explaining our judgements about actions. See Crisp 2006: chs. 1–2 and McElwee 2010b.

blameworthiness, and good deliberation utilize perspectival concepts of reasons or rightness. If they do, we should prefer pluralism. If they do not, we should prefer the simple picture, because it is simpler.

In Chapter 9 I shall argue that, despite what we might first assume, the best utilitarian accounts of praiseworthiness, blameworthiness, and good deliberation do not make use of a perspectival concept of reasons and rightness. On the view I will defend, *non*-perspectival rightness plays a role in explaining these things—though not because being praiseworthy or deliberating well involve trying to discern which of one’s actions is right. Good decision-makers, including those who are virtuous and praiseworthy, have traits that tend to cause them to perform actions that are right in the non-perspectival sense. Non-perspectival rightness is the criterion of good decision-making, even though it is not directly action-guiding. This also means that the non-perspectival concepts of reasons and rightness retain their interest and significance, despite Hooker’s worries.

With those later arguments in mind, I will proceed on the assumption that the simple picture is correct. Rightness is closely connected to reasons, and neither is a perspectival matter. Having said that, it is worth noting that perspectival/non-perspectival pluralists can accept much of what I say in the rest of the book. Since they accept non-perspectival concepts of reasons and rightness, they can interpret my claims as applying to only part of the subject-matter of ethics.

Having assumed that rightness and reasons are closely connected, we might next enquire about the relationship between them. I shall simply assume that rightness is a function of reasons, and that it is to be explained in terms of them: reasons are the sole right-makers.⁴⁰ Explaining rightness is then a matter of identifying what reasons there are, and identifying the function from reasons to rightness.

We can, if we wish, distinguish different kinds of rightness according to the domain of reasons involved: we may say, for example, that prudential rightness is a function of prudential reasons, while moral rightness is a function of moral reasons. Such distinctions may be useful for some purposes. However, utilitarians are quite likely to think that all reasons are, fundamentally, of the same kind—since they are all to be explained in terms

⁴⁰ This entails that there are no ‘deontic reasons’—that is, no reasons provided by the property of wrongness. On deontic reasons see Scanlon 1998: 10–11 and Parfit 2011: 201. Adams 1999: 241 claims that ‘obligations constitute reasons’. I assume to the contrary that obligations are a function of reasons.

of some relationship to well-being.⁴¹ For that reason, utilitarians are quite likely to find the distinction between prudential and moral reasons to be shallow, and to think that for most purposes we should operate instead with an unrestricted concept of rightness, which is a function of all the reasons for action there are.

3.5 Utilitarian Theories of Reasons

If rightness is a function of reasons, the first stage in trying to identify which actions are right, and in trying to explain what makes them right, is to identify what reasons there are. We will take steps in that direction in the chapters to come. But before we get started it is helpful to observe that many different theories of reasons are compatible with the commitments of utilitarianism.

One variable is the theory of well-being a utilitarian theory employs. Different theories of well-being will have different implications about what reasons there are. We shall address the topic of well-being—though without settling on a specific theory of its nature—in the next chapter. In the remainder of this chapter, we will consider some of the other variables that distinguish different utilitarian theories of reasons.

It is common to distinguish between *direct* and *indirect* utilitarian theories of reasons. Act Utilitarianism offers a direct theory of reasons, in the sense that it seeks to explain the reasons for any action in terms of the goodness of the action itself. In contrast, other forms of utilitarianism offer indirect theories of reasons for action.⁴² Rule Utilitarianism seeks to explain the reasons for any action in terms of the goodness of a set of rules that is appropriately related to the action (for example, a set of rules that permits the action). Motive Utilitarianism seeks to explain reasons in terms of the

⁴¹ Admittedly, there are some powerful counterexamples among the great utilitarians. For example, Sidgwick wrote: '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' 1907: 498. Mill 1998/1861: ch. 5 also distinguishes between moral reasons and other considerations.

⁴² In fact, many discussions of utilitarianism are couched as claims about rightness only, and not as claims about reasons. But the theories of reasons implicitly offered by Rule Utilitarianism and Motive Utilitarianism are indirect.

goodness of motives that are appropriately related to the action (for example, a set that causes the action).

Shelly Kagan describes these differences in terms of each theory's 'evaluative focal points'.⁴³ Act Utilitarianism, Rule Utilitarianism, and Motive Utilitarianism all offer theories of reasons for action. So they all, in a sense, 'evaluate' actions—they all make claims about which actions are favoured by reasons. Each therefore includes actions as an 'evaluative focal point'. But they differ in their choice of 'primary' evaluative focal point. Act Utilitarianism evaluates actions in terms of the goodness of actions; Rule Utilitarianism evaluates actions in terms of the goodness of sets of rules; Motive Utilitarianism evaluates actions in terms of the goodness of motives. Each theory's primary evaluative focal point is the thing that, according to the theory, should be evaluated directly by its relationship to good outcomes. Thus the primary evaluative focal point of Act Utilitarianism is the act itself; for Rule Utilitarianism it is the set of rules; and for Motive Utilitarianism it is the motive or set of motives.

Kagan illustrates these ideas with diagrams like those shown in Figure 3.6. For each theory, the item shown in capital letters is the theory's primary evaluative focal point. Act Utilitarianism is 'direct' in the sense that its primary evaluative focal point is identical to the things it evaluates (actions). Rule and Motive Utilitarianism are 'indirect' in the sense that their primary evaluative focal points (rules, or motives) are not identical to the things they evaluate (actions).

However, Kagan goes on to point out that this familiar way of describing things is, in one way, potentially misleading. This is because each theory probably seeks to evaluate other things as well. In particular, Act Utilitarians

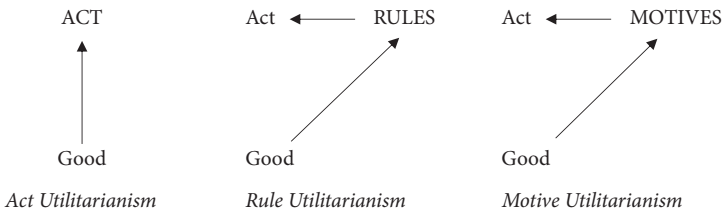


Figure 3.6 Kagan-diagrams of Act, Rule, and Motive Utilitarianism

Source: Adapted from diagrams in Kagan 2000: 136. Reproduced with permission of the Edinburgh University Press Limited through PLSclear.

⁴³ Kagan 2000: 134.



Figure 3.7 Kagan-diagrams of Act Utilitarianism's indirect evaluation of rules and motives

Source: Adapted from diagrams in Kagan 2000: 136. Reproduced with permission of the Edinburgh University Press Limited through PLSclear.

will probably want to say something, as part of their larger theory, about which rules and motives are good. They may say, for example, that a rule is good to the extent that trying to follow it causes agents to perform right actions; or that motives are good to the extent that having them causes agents to perform right actions. These claims will enable Act Utilitarians to make sense of a greater proportion of our ethical lives, and may be particularly important in explaining, for example, how institutions should function or how individuals should take decisions.

The upshot of this, however, is that Act Utilitarianism loses its title as the direct version of utilitarianism. For once we add these other kinds of evaluation, it too appears to be indirect in some respects. This is illustrated in Figure 3.7. Though Act Utilitarianism's evaluation of acts is direct, its evaluation of rules and motives is indirect: it evaluates them in terms of the goodness of the acts to which they are suitably related.⁴⁴

Strictly speaking, we should not call any of these theories 'direct' or 'indirect'. Instead, we can describe each theory's *evaluation of a particular item* as direct or indirect. Act Utilitarianism evaluates actions directly and evaluates rules and motives indirectly, for example, while Rule Utilitarianism evaluates rules directly and evaluates acts and motives indirectly. Kagan's point is that we should keep two issues separate. One is which items (for example, acts, rules, motives, policies, institutions) a theory seeks to evaluate. The other is how it seeks to evaluate each item (directly or indirectly).

There is another important theory in the vicinity, which unfortunately goes by a misleading name. So-called 'Global Consequentialism' evaluates every possible item in a direct consequentialist way.⁴⁵ Thus for example it says that the best sunsets are those with the best consequences, the best acts

⁴⁴ Kagan 2000: 135–6. See also Kagan 1998: ch. 6.

⁴⁵ Pettit and Smith 2000: 121. See also Parfit 1987: 25 and Driver 2014.

are those with the best consequences, the best rules are those with the best consequences, and so on, for every possible object of evaluation. Philip Pettit and Michael Smith write:

[T]he crucial feature of global consequentialism is that it does not privilege any category of evaluand. In particular, it does not privilege the category of acts that has often been privileged, by default, in much consequentialist writing. It does not say, for example, that the right motive-set for someone to have, or the right set of rules for someone to have internalised, is that set which would promote the choice of the right acts.⁴⁶

Now this is an important kind of theory, but the way Pettit and Smith present it is unfortunate. Their presentation yokes together stances on just the two issues that Kagan's discussion helpfully distinguished: which items get evaluated, and how. The name 'Global Consequentialism' suggests that its defining feature is that it evaluates every possible item. Though it has this feature, it also has the feature—identified by Pettit and Smith as 'crucial'—that it evaluates each item directly. A more accurate name for this theory would be, therefore, 'Global Direct Consequentialism'.⁴⁷

In any case, the important thing is that we should not be misled into thinking that there is any necessary connection between extending a theory so that it evaluates every possible item—going global—and settling on a direct form of evaluation for any or all items. The name 'Global Consequentialism' unfortunately obscures the possibility and interest of other consequentialist theories that apply to every possible item, but which offer indirect evaluations of some or all items.

Another unhelpful presupposition afflicts almost all discussions in this area. This is that each item should be evaluated in just *one* way: either directly, or in one specific indirect way. Act Utilitarianism and Global Consequentialism each evaluate acts only directly, in terms of their own consequences. Rule Utilitarianism and Motive Utilitarianism each offer one specific indirect way of evaluating acts. Kagan's diagrams all picture only one relationship between any evaluative focal point and good outcomes: either a direct relationship, or a single indirect relationship; each item is at the tip of only one arrow. This amounts to assuming what I will call *monism* about the evaluation of items. All of the theories we have discussed so far,

⁴⁶ Pettit and Smith 2000: 122.

⁴⁷ Kagan is careful to say that '[a] theory that is direct with regard to *all* the focal points might be usefully labelled *everywhere direct*' 2000: 151, emphasis in the original.

including Global Consequentialism, are monist in this sense. In fact, monism of this sort is very rarely questioned and its presupposition is rarely noticed.

Many monist utilitarian theories of reasons are possible. These include different forms of Act Utilitarianism (which, in its claims about actions, coincides with Global Utilitarianism), plus different forms of Rule and Motive Utilitarianism and other indirect theories of reasons. But we should not confine our attention to monist theories. According to *pluralism* about the evaluation of items, a single item can be evaluated in more than one way. In the case of reasons for action, pluralist theories claim that more than one relationship between an action and good outcomes is a source of reasons for or against the action.⁴⁸ I shall argue in Chapter 5 in favour of a pluralist theory of reasons. But the present point is that the range of possible utilitarian theories of reasons is not restricted to monist theories.

For all of these theories, further questions arise about how to ‘embed’ the primary evaluative focal point.⁴⁹ For example, consider Rule Utilitarianism. Since rules by themselves do not have consequences, in developing a Rule Utilitarian theory we have to specify what we mean by ‘the value of a set of rules’. Does it have to do with the consequences of people complying with the rules, or alternatively with the consequences of people internalizing them? Does it depend on what would happen if everyone were to internalize, or comply with, the rules, or on what would happen if some subset were to do that?⁵⁰ Similar issues arise for Motive Utilitarians. Does the value of a set of motives depend on what would happen if everyone had those motives, or on what would happen if some subset of people had them? Since motives and beliefs are psychologically connected, which beliefs should we imagine these agents having?⁵¹

Less obviously, issues of embedding arise also in the case of Act Utilitarianism. The goodness of an act, we might say, depends just on the consequences it would have in the circumstances. This is the most obvious, and perhaps the best, way to embed actions. But it is worth noting other

⁴⁸ Kagan 2000: 153 mentions the possibility of pluralism in passing. Mulgan 2001 presents a pluralist theory, as does Woodard 2008a. Note that this kind of pluralism is not to be confused with perspectival/non-perspectival pluralism.

⁴⁹ Once again this is Kagan’s helpful terminology. See Kagan 2000 and Kagan 1998: 227–30.

⁵⁰ There has been considerable discussion of these issues recently. See for example Hooker 2000: 80–5, Mulgan 2001: ch. 3, Mulgan 2006: ch. 5, Ridge 2006, Smith 2010a, Tobia 2013, Toppinen 2016, Mulgan 2017, Yeo 2017, and Podgorski 2018.

⁵¹ Bradley 2018 is a helpful discussion of embedding issues as they arise for consequentialist theories of virtue.

possibilities. One way to think of perspectival forms of Act Utilitarianism is as employing a different way of embedding actions: they specify the value of actions in terms of the agent's beliefs, or what it would be reasonable to believe in the circumstances; the value of the action is then taken to be the value it would have if embedded in circumstances corresponding to the content of these beliefs.⁵² Another kind of embedding for Act Utilitarianism is given by 'possibilists', who specify the value of an action not in terms of the consequences it would have if it were performed in the circumstances, but in terms of the consequences it would have if it were performed and the agent were to do everything else that she ought to do.⁵³

As these brief remarks suggest, each possible direct, indirect, or pluralist utilitarian theory of reasons can be subdivided into different versions, according to how the theory embeds the items it utilizes in its explanation of reasons. Some of these possible versions of each theory will be very implausible for one reason or another, and some may be equivalent to each other. But it is likely that many possible utilitarian theories of reasons are worth exploring.

3.6 Conclusion

Our approach will be to develop a version of utilitarianism with a theory of non-perspectival reasons at its core. Using the simple picture discussed in Section 3.4, we will assume that reasons and rightness are closely connected to each other, but that neither is closely connected to praiseworthiness, blameworthiness, or good deliberation. We will further assume that the rightness of actions is a function of reasons. In explaining what reasons there are, we hope to make some progress toward explaining the rightness of actions. On the other hand, we will have to give separate explanations of other important parts of ethics, including the nature of good decision-making and the grounds of praiseworthiness and blameworthiness.

⁵² Kagan 2000: 145.

⁵³ See Zimmerman 1996: ch. 6 and Woodard 2009.

4

Well-Being

Utilitarians claim that well-being is the sole fundamental good, and that we can explain reasons, rightness, justice, and virtue in terms of it. For this reason, utilitarians have a strong interest in knowing what well-being is.

Traditional philosophical theories of well-being aim to specify the nature of well-being. That is, they attempt to identify its constituents and explain why it is that these things are its constituents. For example, hedonists claim that the constituent parts of well-being are pleasures, while an ‘objective list’ theory may claim that autonomy, say, is a constituent of well-being.¹ Desire theorists claim that what makes something a constituent of someone’s well-being is that she desires it, or that she would desire it under the right circumstances. Perfectionists claim that what makes something a constituent of well-being is that it is in some way excellent.²

Several of these different theories about the nature of well-being are compatible with utilitarianism. Moreover, we can give utilitarianism a somewhat different complexion by adopting one theory rather than another. Thus it is undoubtedly important, for utilitarians, to try to resolve these traditional philosophical disputes about well-being. Unfortunately, however, these disputes are very difficult to resolve. Arguably, the central question for philosophical theories of well-being is whether some form of subjectivism is true—that is, whether what makes something good for someone is entirely a matter of whether she would value it in some specified circumstances. In my view, however, the best existing arguments both for and against subjectivism are inconclusive. So far as that central issue goes, I believe, we should currently remain agnostic.

However, it does not follow that we do not know anything about well-being, nor that we do not know enough to articulate a plausible form of

¹ On hedonism see Feldman 2004 and Crisp 2006: ch. 4. On Objective List theories see Fletcher 2013, Rice 2013, Bradley 2014, and Hooker 2015.

² On the desire theory see Griffin 1986, Arneson 1999a, M. C. Murphy 1999, Heathwood 2006, and Heathwood 2016. On perfectionism see Hurka 1993 and Kraut 2007.

utilitarianism. Alongside their theoretical interest in the nature of well-being, utilitarians have a practical interest in understanding how to promote well-being. And it may be that we can make some progress with this question before we have settled traditional philosophical debates about the nature of well-being. Moreover, the question of how to promote well-being is not merely of practical interest for utilitarians. The plausibility of their theory depends on its implications about particular cases, and these implications depend in turn on what would promote well-being.

In this chapter I shall first examine the traditional debates about the nature of well-being, concluding that these debates are currently in stalemate. I shall then shift focus from the metaphysics to the epistemology of well-being. I shall claim that we can learn an important general truth about how to promote well-being from considering the circumstances under which someone knows what is good for her.

4.1 Philosophical Theories of Well-Being

Well-being is the value of a person's life for the person whose life it is. A life can have other kinds of value. For example, it might be beautiful, or morally good, or instructive. Among the kinds of value that lives can have, well-being is that kind which is good or bad for the person whose life it is.³

The concept of something's being 'good for' someone can seem mysterious.⁴ It is not the same as the concept of something's being good *according to* someone. Many of the things that I take to be good have nothing to do with me, and for that reason they are not good for me. Other things that I take to be good are not really good at all, and for that different reason they are not good for me. Nor is what is good for me simply the intersection of what is good and what is mine. A work of art does not become a constituent of my well-being merely by coming into my possession.⁵

³ 'Well-being' has a narrow and a broad sense. In the narrow sense it is necessarily positive, and its contrary is 'ill-being'; in the broad sense it includes ill-being, and a person's overall level of well-being may be negative. In what follows I use 'well-being' in the broad sense. On possible asymmetries between well-being in the narrow sense and ill-being see Kagan 2014.

⁴ G. E. Moore 1993/1903: 150 claimed to be able to understand 'good', but not an independent kind of 'goodness for'. As we noted in Chapter 3, Judith Jarvis Thomson claimed to be able to understand 'good for', but not 'good' 1997: 276. It is possible to find either concept puzzling, but neither is so mysterious that we cannot use it.

⁵ This objection to the possessive account, and the example, are due to Fletcher 2012: 3. Fletcher goes on to defend a 'locative' analysis of the concept of well-being. Darwall 2002 and Campbell 2013 offer alternative analyses.

However, I take it that we have a sufficient grasp of the concept of something's being good for someone, whether or not we can also give an analysis of it. It is the concept at play when we seek to benefit or harm someone, including ourselves. To benefit someone is to bestow something that is good for him, not merely good. Obviously enough, it can be very difficult to know what would benefit or harm someone, and attempts to do these things can easily misfire. But that is a different problem: it is the problem of knowing which things are or would be good or bad for someone, and our sense that this is a genuine problem rests on our grasp of the concept of well-being. Similarly, we think there is a genuine problem in deciding which objects are beautiful and which are not in part because we have a sufficient grasp of the concept of beauty.

A person's well-being depends, presumably, on what his life contains. Among the things we might expect to be relevant are his achievements, relationships, sorrows, and pleasures. Each of these things may contribute to his well-being in one or both of the following ways. First, it may contribute instrumentally, by causing or preventing something else that is good or bad for him. Second, it may contribute noninstrumentally, by being good or bad for him independently of its causal consequences.

We can define a 'constituent' of a person's well-being as anything that is noninstrumentally good or bad for him. A person's well-being depends only on the value of these constituents. The instrumental value of the other parts of his life is entirely accounted for by the noninstrumental goods and bads they cause or prevent.⁶

These remarks suggest an agenda for philosophical theories of well-being. For any given entity with well-being, we would like to know the constituents of its well-being. Ideally, we would also like to know why these things are the constituents of its well-being—both for the sake of deeper understanding, and so that we understand differences between subjects and circumstances. As it is sometimes put, we would like to *enumerate* the constituents of each subject's well-being, and also to *explain* why it is that these things are constituents of that subject's well-being. Philosophical theories of well-being typically try to answer these two questions.⁷ To answer them would be to give, as I shall say, an account of the nature of well-being.

⁶ A further issue is whether well-being is simply the sum of the value of the constituents of well-being, or also depends on the 'shape of a life'. See Velleman 1991 and Glasgow 2013. It may be possible for utilitarians to apply sum-ranking across persons yet not within lives. However, I shall assume that it is applied both across and within lives. See Chapter 2 Section 2.4.

⁷ Crisp 2006: 102–3. Crisp here uses a broader sense of 'constituent' than I have adopted. For example, he allows that accomplishment may be a 'constituent' of well-being according to

Hedonists, for example, claim that the only constituents of well-being are pleasures and pains. Desire theorists claim that the only constituents of well-being are satisfaction or frustration of (some specified set of) the subject's desires.⁸ 'Objective List Theories' specify some list of goods—such as, for example, pleasure, friendship, achievement, knowledge, and autonomy⁹—and claim that the constituents of well-being are the ones on the list.

These are rival answers to the first question about the nature of well-being, which asks us to enumerate the constituents of well-being. These theories pair up with associated answers to the second question, which asks in virtue of what these things are constituents. For example, hedonists may answer the second question by saying that pleasures are constituents in virtue of their phenomenological quality; desire theorists may answer by saying that an object which satisfies a desire is a constituent in virtue of its satisfying a desire. These pairings sometimes seem trivial or uninformative, but they are sometimes surprising. For example, some hedonists may claim that pleasures are constituents of well-being because they consist in the satisfaction of desires.¹⁰ In principle, each enumeration of constituents could be paired with more than one answer to the explanatory question.

On the other hand, each answer to the explanatory question, taken together with all relevant facts about a subject, must yield exactly one enumeration of constituents for that subject. In this sense, the explanatory question is the most fundamental question about the nature of well-being. If we manage to answer it, we get an enumeration and a unique explanation of why the things enumerated are constituents of well-being.

Unfortunately, we are not yet in a position to answer it. The most important debate between explanatory theories is between subjectivist and non-subjectivist theories. Subjectivists claim that the answer to the explanatory question is to be found in some specification of the subject's *valuing attitudes*. In its most ambitious form, subjectivism is a claim about the well-being of any subject:

hedonism, whereas I have stipulated that constituents are noninstrumentally good or bad for the subject. Lin 2017b claims that all major theories of well-being answer both the enumerative and the explanatory question.

⁸ This is the simplest form of desire theory. According to what Bradley 2014: 235 calls the 'combo view', what is good for someone is the combination of desiring something and getting it. I will ignore this alternative.

⁹ This is the list proposed in Hooker 2015.

¹⁰ Heathwood 2006: 555–9. Note that Heathwood defines 'desire satisfaction' non-standardly, as belief that the object of the desire obtains (548).

Subjectivism For any subject *S*, and any thing *X*, *X* is or would be a constituent of the well-being of *S* if and only if, and because, in the right circumstances *S* would value *X* noninstrumentally.¹¹

Different versions of subjectivism specify the relevant valuing attitudes, and ‘the right circumstances’, differently. For example, one version might characterize valuing as desiring, while another characterizes it as believing good; and one might characterize the right circumstances as the subject’s actual circumstances, while another characterizes them as ones of full information.¹²

In any of its forms, subjectivism is a psychologistic theory of value: it seeks to explain a kind of value entirely in terms of psychological facts. Together with its generality, this makes it an extremely ambitious theory. It also explains one source of its appeal. Since it seeks to explain well-being entirely in terms of psychological facts, it seems to be metaphysically modest.¹³

Subjectivism is also, unsurprisingly, a subject-centred theory of value. According to subjectivism what is noninstrumentally good for me is not determined by just anyone’s attitudes—for example, by some guru’s—but by my own attitudes.¹⁴ This is the second source of subjectivism’s appeal. Because it is subject-centred, it meets a constraint of subjective endorsement that many have taken to apply to theories of well-being. Peter Railton gave an influential statement of this idea, as follows:

[W]hat is intrinsically valuable for a person must have a connection with what he would in some degree find compelling or attractive, at least if he were rational and aware. It would be an intolerably alienated conception of someone’s good to imagine that it might fail in any such way to engage him.¹⁵

This is an attractive idea. As we noted earlier, the idea that something is noninstrumentally good for someone is not equivalent to the idea that it is noninstrumentally good according to him. But it is hard to believe that what is noninstrumentally good for someone is not related at all to what he values or finds engaging.¹⁶

¹¹ This definition is indebted to Dorsey 2012: 407. See also Lewis 1989: 113.

¹² See Dorsey 2012.

¹³ See Lewis 1989: 113. How modest it is depends on the metaphysics of the relevant psychological facts.

¹⁴ Kraut 1994: 42–3. ¹⁵ Railton 2003b: 47.

¹⁶ See Rosati 1996, Sumner 1996: 42, and Dorsey 2017. For criticism see Sarch 2011.

However, neither of the attractive features of subjectivism provides a compelling argument for it. Its metaphysical modesty fails to do so for two reasons. The first is that metaphysical modesty is just one desideratum for a theory of the nature of well-being; in principle, we should be willing to trade it off against other theoretical virtues. More importantly, non-subjectivist theories can be equally metaphysically modest. Some forms of hedonism, for example, are not subjectivist, since they deny that subjective approval of pleasure is necessary for pleasure to be a constituent of well-being. But if they claim that what makes pleasure a constituent of well-being is its phenomenological quality, they are not obviously more strongly metaphysically committed than are subjectivist theories.¹⁷

Nor does compatibility with Railton's anti-alienation constraint provide a good argument for subjectivism. This is obvious if we state Railton's idea in a form similar to our earlier statement of subjectivism:

Railton For any subject *S*, and any thing *X*, *X* is or would be a constituent of the well-being of *S* only if in the right circumstances *S* would value *X* noninstrumentally.

There are two significant differences between this claim and the claim made by subjectivism. First, subjectivism claims that there is a bi-conditional relationship between being a constituent and being valued in the right circumstances ('if and only if'), while this formulation of Railton's idea claims only that there is a one-way conditional relationship ('only if'). Railton's idea allows, while subjectivism denies, that something could be valued noninstrumentally by the subject in the right circumstances and yet fail to be a constituent of her well-being.

This is important because it means that Railton's idea is compatible with other necessary conditions for something's being a constituent of well-being. For example, according to some 'hybrid' theories there are two necessary conditions for something's being a constituent of well-being: one is that the subject endorses it in some way, and the other is that it is objectively valuable. These theories are compatible with Railton's idea, but they are incompatible with subjectivism.¹⁸

The second difference between the two claims is just as important. Subjectivism claims that something is a constituent of well-being *because*

¹⁷ Hooker 1991 offers further criticism of the argument for subjectivism from metaphysical modesty.

¹⁸ On hybrid theories see Kagan 2009, Sarch 2012, and Woodard 2016.

it is valued, while Railton's idea claims only that when something is a constituent of well-being it is also valued. It does not commit itself to any explanation of this correlation. Railton's idea is thus compatible with the anti-subjectivist idea that what people value, or would value, is partly or wholly determined by what is valuable.¹⁹

In my view, the arguments against subjectivism are no more compelling. One argument is that subjectivist theories are extensionally inadequate: that is, they have implications about which things are constituents of well-being, or why, that fail to match our intuitions sufficiently well. For example, if someone values smashing icicles for its own sake, subjectivism seems to imply that smashing icicles would be noninstrumentally good for her.²⁰

One reply to this sort of criticism is that subjectivist theories are, in fact, able to match the relevant intuitions reasonably well. In this connection it is important to note that subjectivism is perfectly compatible with the idea that subjects can make mistakes about what is good for them. One reason for this is that subjects can overlook the instrumental disvalue of satisfying their desires.²¹ Another is that subjects might not have good epistemic access to the relevant valuing attitudes—particularly in cases where their values change over time, as we will discuss in Section 4.4. Many alleged counterexamples to subjectivism fail for one or both of these reasons. For example, the person who values smashing icicles may be making a prudential mistake in setting about smashing them, even by subjectivism's lights. Doing so may get in the way of satisfying too many of her other values. It may also prevent her from developing new interests that would yield more satisfaction over the course of her life.

A natural reply for the critic of subjectivism is to shift attention from the judgement that smashing icicles is overall bad for this subject to the explanation of why it is bad. It is bad, the critic claims, not merely because of its admitted instrumental disvalue, but also because it has no positive noninstrumental value. This is a tempting thought, but I doubt that it is a sufficient basis to reject subjectivist theories. For it to have that significance we would need to have confidence in our ability to sort things into the following two categories: those which are reliable causes of well-being (or ill-being), and

¹⁹ Hurley 1989: Part 1 claims that the relationship between preference and value is one of interdependence. For further discussion of the anti-alienation constraint (sometimes called 'internalism'), see Arneson 1999a: 135–41 and the excellent discussion in Fletcher 2016: 65–75.

²⁰ Kraut 1994: 42.

²¹ Heathwood 2005.

those which are noninstrumentally good (or bad). But we should not have much confidence in our ability to do this. Thought experiments designed to bring out these intuitions typically involve extremely unusual cases, and our intuitions about these cases may not have much credibility. At least some of our judgements that something is a constituent of well-being may reflect common wisdom about general causal relationships, rather than insight into the metaphysics of well-being.

A second argument against subjectivist theories appeals to a highly abstract claim about the relationship between value and valuing. According to this argument, valuing is normatively significant only to the extent that it is appropriate, and it is appropriate only when its object is valuable. Thus, in determining what is good for people, objective value is more basic than valuing.²² In reply, the subjectivist can again point out that she is able to explain some intuitions in the vicinity of this claim. For example, she can explain the judgement that it would be better, for the subject herself, to have some desires rather than others. She can also point out that it is compatible with subjectivism to claim that when we desire something we take it to be *desirable*; its desirability is explained, according to the subjectivist, by the features that sustain our desire in it. To the extent that these two points fail to capture the thesis that value is prior to valuing, the subjectivist can deny that we have good reason to trust intuitions in favour of that thesis.²³ It is a highly abstract claim about the metaphysics of value, and it is not clear that we have much intuitive insight into such matters.²⁴

These remarks about the arguments for and against subjectivism are, of course, extremely compressed. There is much more to be said about these and other arguments, and I do not pretend to have shown that none of the arguments is compelling.²⁵ However, they may serve to convey why it seems to me that these arguments currently result in stalemate.

²² Kraut 1994: 43–5 makes a version of this argument.

²³ Griffin 1986: 26–31 makes similar points about the thesis that value is prior to desire.

²⁴ Hurley 1989: Part I argues that subjectivism about value is incoherent, because preferences are individuated partly by their relations to objective values. This is a more interesting argument than the bare appeal to metaphysical intuition about the relationship between value and valuing. However, Hurley's argument depends on controversial theses in the philosophy of mind and language. Lillehammer 2007 is one example of critical discussion.

²⁵ Sobel 2016 is a recent defence of subjectivism. Dorsey 2013 and Bradley 2016 explore difficulties for subjectivists in giving an account of the relationship between lifetime well-being and well-being at a time. Lin 2017a argues that subjectivist theories fail to give a plausible account of children's well-being.

4.2 What We Know about Well-Being

This is a disappointing conclusion. However, we should not infer that well-being is simply a mystery. In fact, we know quite a bit about it. First, we have a reasonably good grasp of the concept of well-being itself, as I have mentioned. This is the concept of the value of a person's life for her, and it is closely connected to the concepts of harm, benefit, and self-interest. When welfarists claim that the goodness of outcomes depends entirely on the well-being they contain, we have a reasonably good idea of what they mean.

Second, we are quite confident in making some judgements about levels of well-being. If a relative of yours has recently had serious health problems, you may be very confident that her well-being has taken a hit. You do not need to know whether her very poor health is a constituent of ill-being, or merely a cause of it, to be confident in making this judgement. Similarly, you may judge quite confidently that your friend is doing much better since he changed jobs.

Everyday judgements like these are not very sensitive to the differences between philosophical theories of well-being.²⁶ Hedonists, Desire Theorists, Objective List Theorists and others are likely to agree with common opinion across a broad range of cases about many of the things that are good and bad for people. For example, these theories may all agree that autonomy, pleasure, health, friendship, and education are usually good for people, and that suffering, social isolation, depression, and economic insecurity are usually bad for people. Their disagreement will be about which of these things are merely instrumentally good or bad and which are constituents of well-being or ill-being. However, for many purposes that distinction does not matter. Indeed, for some purposes that distinction does more harm than good. A common mistake in thinking about well-being is to assume that if something is a constituent of (positive) well-being, it must make a positive contribution to well-being overall. This overlooks the prudentially very important fact that its noninstrumental value may be smaller than its instrumental disvalue. Something that is a constituent of your (positive) well-being may easily be bad for you overall.²⁷

Because they often depend on facts about causes, these generalizations about the sources of well-being need to be hedged. In particular, they may be

²⁶ Compare Hausman 2012: 92: 'Economists and everyday folk do not have to wait for a satisfactory philosophical theory of welfare before they can say anything about what makes people's lives better or worse.'

²⁷ Heathwood 2006: 546–7.

true only within some contexts or circumstances. The contribution of autonomy to well-being may depend on cultural context and on personal tastes and preferences, for example. Still, we know something about these generalizations, including about their limitations. Moreover, we can expect to learn more about them, and about cultural variations, from the currently burgeoning ‘science of well-being’.²⁸

These generalizations are especially helpful for public policy and other kinds of centralized beneficence. Policymakers need to know what is generally beneficial more than they need to know what suits any particular individual. Generalizations about well-being may also be of some help to individuals trying to promote their own well-being or that of someone close to them. Knowing that unemployment is correlated with a long-lasting deficit in self-reported well-being, for example, may help to inform someone’s decisions.²⁹

However, we want to know more than these generalizations tell us about how decisions will affect us as individuals. They do not track all of the differences between individuals in tastes, ambitions, temperament, and circumstances. When deciding what career to pursue, whether to start or end a relationship, or whether to have children, we want to know how these things will affect us in all of our individuality. The same goes for less dramatic decisions.

Any account of the state of our knowledge of these matters should be consistent with the fact that it is difficult to be prudent. We often find purely prudential decisions agonizing, and we often seem to get them wrong. Though we may not regret a large proportion of our decisions, this may be because we adapt and learn to live with their results, or because we engage in post hoc rationalization of poor decisions.³⁰

One element of the difficulty of prudence is the difficulty of predicting the consequences of our various options. But even when we know reasonably well what the consequences of our various options will be, prudence can be difficult because we do not know which outcome would suit us best. On top of the difficulties of predicting outcomes, we have the difficulties of evaluating them. Similar remarks go for beneficence directed at those we

²⁸ See Kahneman et al. 1999 and Huppert et al. 2005. For discussion of the relationship between the philosophy and the science of well-being see Alexandrova 2016 and Wren-Lewis 2014. On the significance for philosophical discussions of well-being of psychological research about cultural differences see Tiberius 2004.

²⁹ Lucas 2007.

³⁰ Lucas 2007.

know well. For example, it can be difficult to know how a well-intended act will be received.

One of the main ways that well-being gets promoted is by people trying to promote their own well-being and the well-being of those close to them. They are hampered in doing so by the difficulty of prudence and beneficence, which in part is a matter of the difficulty of knowing what would be good for the person concerned. An important question for utilitarians, therefore, is under what circumstances people could know this more easily. If we could somehow make it easier to know, we would boost people's power to promote their own well-being and that of others.

4.3 Alienation as Evidence

Consider again Railton's anti-alienation constraint. It is, I claimed, an attractive idea, but it has the wrong form to be used in arguments for subjectivism. One problem was that it asserts a conditional rather than a bi-conditional relationship between something's being a constituent of well-being and its being valued in the right circumstances. This means that it is compatible with additional necessary conditions on something's being a constituent of well-being, such as those proposed in (non-subjectivist) hybrid theories. The other problem was that it asserts a correlation, rather than an explanatory relationship, between being valued in the right circumstances and being a constituent of well-being. This means that it is compatible with the anti-subjectivist idea that what people value is in part determined by what is valuable.

However, we can use Railton's idea for a different purpose. Rather than trying to derive conclusions about the nature of well-being from it, we can use it to learn something about how we can know what is good for us. The correlation proposed by Railton's idea suggests that, if a subject would not value X in the right circumstances, X is not or would not be a constituent of that subject's well-being. If that is correct, we can treat not valuing X in the right circumstances as *evidence* that X is not (or would not be) a constituent of the subject's well-being. The correlation is sufficient for evidence, though it is not sufficient to draw conclusions about the metaphysics of well-being.

Less obviously, we can treat valuing X in the right circumstances as evidence that X is (or would be) a constituent of S 's well-being. This is less obvious, because we might at first think that Railton's idea has the wrong form to provide positive evidence of this sort. As we have noted, it asserts a

conditional relationship, not a bi-conditional relationship, between valuing and being a constituent of well-being: it claims that *X* is a constituent only if it would be valued in the right circumstances. This claim is compatible with the existence of additional necessary conditions for *X*'s being a constituent of well-being, including for example the requirement that *X* is objectively valuable.

However, learning that *S* would value *X* in the right circumstances is evidence of a sort that it is (or would be) a constituent of *S*'s well-being. It is evidence in the sense that it raises the probability that *X* is a constituent, relative to a state of ignorance about whether *X* would or would not be valued. It tells us that *X* is, as it were, in the right ballpark. Since there might be additional necessary conditions on *X*'s being a constituent of *S*'s well-being, this evidence is not conclusive, and we might get false positives.³¹ However, that is true of many diagnostic tests.

Consider the following analogy. Suppose that we want to know which things are truly funny, and we believe that something cannot be truly funny if, in the right circumstances, some specific person (an average person, say, or perhaps an expert judge) would not laugh at it. Suppose also that we are unsure whether there are additional necessary conditions on something's being truly funny, such as satisfying some objective comedic principles. Now, if we find out that the relevant person would not laugh at *Y* in the right circumstances, then we have strong evidence that *Y* is not truly funny. But we also learn something if we find that, in the right circumstances, this person *would* laugh at *Z*. We have learned that *Z* meets one of the necessary conditions for being truly funny, and this raises its probability of being truly funny relative to a state of ignorance about whether it meets that condition. To be sure, our negative evidence about *Y* is stronger than our positive evidence about *Z*—but we have some evidence that *Z* is truly funny, nonetheless.

For the same reason, we can use Railton's idea for both positive and negative epistemic purposes. If, in the right circumstances, a subject would not value something, that is strong evidence that this thing is not, or would not be, a constituent of her well-being. On the other hand, if, in the right circumstances, she would value it, this is weaker evidence that it is or would

³¹ If we thought that there are additional necessary conditions that are highly restrictive, we might not treat the fact that *S* would value *X* in the right circumstances as evidence that *X* is or would be a constituent of *S*'s well-being, in some stronger sense of 'evidence'. However, even in that circumstance we should accept that it would be evidence in a weaker, probability-raising, sense.

be a constituent of her well-being.³² I shall call these uses of Railton's idea 'the alienation test'. Obviously, to interpret this test more precisely we need to understand what is meant by 'valuing' and 'the right circumstances'.

The leading subjective theory of well-being, the desire theory, identifies 'valuing' with desiring. According to this theory, to value something non-instrumentally is to desire it noninstrumentally. However, if desire is understood narrowly as a motivational state disposing the agent to act, it seems likely that we can value something without desiring it.³³ Motivation is a scarce resource, and we must ration it. In particular, we must ration it according to judgements of feasibility. This is why most people grow out of the desire to become an astronaut, even if they continue to think that it would be great to be an astronaut.³⁴

As we broaden the notion of desire the claim that valuing is desiring becomes more plausible, but it also becomes less informative. In any case, we should be cautious about assuming that to value something is to desire it in the narrow sense. It seems quite likely that valuing can find expression in a range of mental states. Candidates include desiring, believing to be good, hoping and wishing, and affective states such as moods and emotions.³⁵ All of these are in some sense evaluations, and they have a *prima facie* claim to being evaluations in the sense at issue in the alienation test. Suppose, for example, that someone claims that watching opera would be good for me. I do not desire to watch it and I do not believe that watching it would be good for me; neither do I hope nor wish to watch it. Whether or not it

³² To play these evidential roles we do not need Railton's correlation to be exceptionless. If we wished, we could weaken the correlation it asserts and so boost its plausibility. However, I shall ignore that possibility here.

³³ It may also be possible to desire something noninstrumentally without valuing it. Velleman writes: 'If Satan ever loses sight of the evil in what he now desires, if he ever comes to think of what he desires as really good, he will no longer be at all satanic; he'll be just another well-intentioned fool. The ruler of Hell doesn't desire what he wrongly thinks is worthy of approval; he desires what he rightly thinks isn't' 1992: 18. This is of course controversial: Orsi 2015 and Tenenbaum forthcoming are useful surveys. See also Watson 1975: 210 and Railton 2012: 28, where Railton characterizes desire as 'a beckoning idea that draws us forward, rather than a blank drive that pushes us from behind'.

³⁴ Sidgwick wrote: 'a prudent man is accustomed to suppress, with more or less success, desires for what he regards as out of his power to attain by voluntary action... but any success he may have in diminishing the actual intensity of such desires has no effect in leading him to judge the objects desired less "good"' 1907: 110. Velleman 1992: 17 claims that the distinctiveness of desire amongst other conative attitudes (such as wishing or hoping) is that desire aims constitutively at the attainable.

³⁵ See Watson 1975, Smith 1992, Haybron 2008, Hawkins 2010, Dorsey 2012, and Kubala 2017. On the sense in which emotions are evaluations see Deonna and Teroni 2014. As they note (17 n. 3), some ways of thinking of emotions as evaluations may be incompatible with some forms of subjectivism about value.

‘engages’ me, in the sense that makes Railton’s claim most plausible, may then depend on whether it stirs my emotions in some appropriate way.³⁶

One interesting implication of this open-mindedness about valuing is that someone may value a single aspect of a single thing at a single time in divergent ways. There is no guarantee that our desires, beliefs about goodness, hopes, wishes, and affective reactions will always be harmonious in even this quite minimal way. For example, someone may desire close social interaction noninstrumentally, but find such interaction deeply unsettling and agitating. It need not be the case that one or other of these evaluations is inauthentic, or not really his. He may simply have incoherent evaluations of social interaction.³⁷ If that is possible, then there may be no clear answer about what he values most or about what would be best for him.³⁸

Despite this unsettling implication, let us cast the net wide and suppose that any of these states may be a valuing state in the sense relevant to the alienation test. The next question is how to understand ‘the right circumstances’. There are two main aspects to this question: how to handle evaluations based on false beliefs, and how to handle changes in values. Let us briefly consider errors here, before considering changing values in Section 4.4.

Suppose that Anka is a teacher. Unfortunately, she falsely believes that she is a bad teacher. For this reason she finds teaching completely unrewarding and she wants to change career. In fact, she is an excellent teacher, and if only she knew this she would find all of the hard work worthwhile. What does proper application of the alienation test imply about her situation?

The issue is whether we should distinguish between informed and uninformed valuing attitudes when applying the test. Most subjectivists about

³⁶ It is not straightforward to incorporate moods and emotions among the valuing states relevant to Railton’s idea. The difficulty concerns whether we can make sense of valuing something ‘noninstrumentally’ in cases where the valuing state is a mood or emotion. I will not pursue this issue here. If we cannot make sense of this, it may be necessary to gloss Railton’s idea differently or to give up on the idea that moods and emotions can be valuing states in the relevant sense.

³⁷ On inner conflict, see Hurley 1989: ch. 8. See also Hawkins 2010 on conflict between affective and conative evaluations. Wilson et al. 2000 discusses psychological evidence for the possibility of divergent implicit and explicit evaluations of the same object.

³⁸ In such cases there is something to be said for trying to achieve harmony. Other things equal, having harmonious values is likely to make it easier to achieve well-being. But it does not follow that the best life for someone will always be one in which he has harmonious values. Achieving harmony may sacrifice too much. Tiberius 2008: 116 emphasizes the interpretative and creative nature of constructing a coherent self-conception. Hayward 2017 discusses rationality and inner disharmony.

well-being attach great weight to this distinction.³⁹ David Sobel explains one rationale for this. Discussing uninformed desires, he says:

The desire is not responsive to the true nature of its object, and in that sense, is not really for it. It is this distinction, I take it, that we are after when we say that we did not ‘really’ want something. It was only wanted to the extent that we were ignorant of its true nature. Our informed desires are, in a sense, more genuinely for their object.⁴⁰

Applying Sobel’s rationale, we might say that Anka’s desire to stop being a teacher is not genuinely for stopping, because it is based on the false belief that she is a bad teacher.

Some worry that if we seek to correct for uninformed desires we might end up with a picture of a person’s values that is alienating, irrelevant, or indeterminate. Adding or subtracting beliefs from someone’s psyche might have a large effect on her values. If we ask what Anka would value were she well informed, we might end up with a picture of what her counterfactual counterpart, Informed Anka, values, rather than a picture of what Anka herself values. Alternatively, there may be no fact of the matter about what Informed Anka values. To the extent that there is a fact of the matter about it, we may worry that it may have nothing much to do with what would be good for Anka herself.⁴¹

However, given our epistemic aims we can avoid these worries. Rather than attempting to correct for error, we can be more conservative and simply discount any valuing attitudes based on error. We do not have to reach a view one way or the other about the evidential significance of what Informed Anka would value. We can simply say that, so far as the alienation test goes, we require that valuing attitudes are not based on false beliefs about their objects. Since many of our valuing attitudes are not based on false beliefs, sticking to this conservative version of the test will often yield results.

4.4 Changing Values

Now consider the second important issue in the specification of the ‘right circumstances’. Suppose that Sally is wondering whether to give up her job

³⁹ For example, see Sidgwick 1907: 110–11, Brandt 1979: 268, Railton 2003b: 54. M. C. Murphy 1999 is an exception.

⁴⁰ Sobel 2009: 347. Sobel is responding to the claim made by Enoch 2005 that subjectivists have offered no credible rationale for idealization.

⁴¹ Rosati 1995. See also Dorsey 2017.

in the city in order to become a farmer. She likes a lot of things about city life, including the ability to socialize with others and the easy access to many amenities. On the other hand, she grew up in the country and enjoyed its quiet solitude and fresh air, and she knows herself to be very adaptable. She understands that her tastes change, and that many of the things she currently likes about the city would lose their appeal after a while if she were to move to the country.

The question is whether we should apply the alienation test using Sally's current attitudes to life in the country, or using those she would have were she to move there. We can call the first kind *actual-life* evaluation, and the second kind *in-life* evaluation. For this purpose, we can think of Sally as having a different 'life' in the country than she would in the city—not because of the change in her location, but because the move would cause a significant change in her values.⁴²

Sally's choice is represented in Figure 4.1.

In this figure, the fact that Sally's values would not change if she were to stay in the city is represented by the fact that the line connecting her to that future is solid. Her current city-dwelling values (V_c) would persist if she were to make this choice. In contrast, the fact that Sally's values would change significantly if she were to become a farmer is represented by the fact that the line connecting her to that future is dashed. If she were to take this option she would acquire new, farming values (V_f). The question is whether proper application of the alienation test uses her current values for both futures, or whether it evaluates each future by reference to the values that she would have in that future.

In-life evaluation seems clearly superior on intuitive grounds. Actual-life evaluation privileges the subject's existing attitudes, and uses them to evaluate her life even when her values change. Given that Sally's values would change significantly were she to move to the country, it seems perverse to insist that her current attitudes provide good evidence about how well she

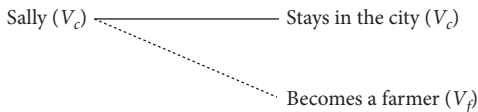


Figure 4.1 Sally's first choice

⁴² On choices involving changing values see Bykvist 2006, Ullmann-Margalit 2006, Harman 2009, Paul 2014, and Pettigrew 2015.

would do under those conditions. It is perverse in roughly the same way that it would be perverse to evaluate her life using someone else's standards.

Admittedly, in-life evaluation might be incompatible with some conceptions of the importance of personal commitments. If our measure of well-being adopts in-life evaluation, prudence seems to favour making transforming choices whenever they maximize well-being. But this would mean that prudence favours, roughly, adopting whichever goals or commitments would result in the best future for you in terms of well-being.⁴³ Suppose, for example, that Sally has a strong personal commitment to working with a certain charity, and that she can fulfil this commitment only if she lives in the city. If she were to move to the country, let us suppose, this commitment would wither away, and all things considered her well-being would be higher. But, it might be thought, her existing commitment should not be treated as dispensable in the way that, it seems, in-life evaluation treats it. It should be privileged in just the way that, a moment ago, I claimed was perverse.

However, we have to be careful to separate different issues. It could be that Sally incurs various obligations in the course of her charity work that limit her moral freedom. The charity might have come to depend on her and it might have acquired a reasonable expectation that she would continue her work with them. This might generate a duty for her not to let them down. Commitments can certainly have that sort of force, and with this in mind it is right to privilege the actual commitments of subjects over counterfactual ones. But the present issue is not about that sort of significance of commitments. Instead we are wondering whether, insofar as someone's future well-being is correlated with her evaluative standards, the correlation is with her existing standards or with the standards she would have under the circumstance in question. So far as that question goes, I claim, we should not privilege the subject's actual commitments.

A second case helps to make the point. Suppose now that Sally must choose between moving to the country to become a farmer, or staying in the city *and* becoming a parent. The special feature of this second case is that every option she faces would cause a significant change in her values. This case is represented in Figure 4.2.

⁴³ I am here using 'prudence' to mean doing whatever maximizes personal well-being, not to refer to a kind of rationality. As McKerlie 2013: ch. 7 notes, change in values may have different significance for well-being than it has for rationality.

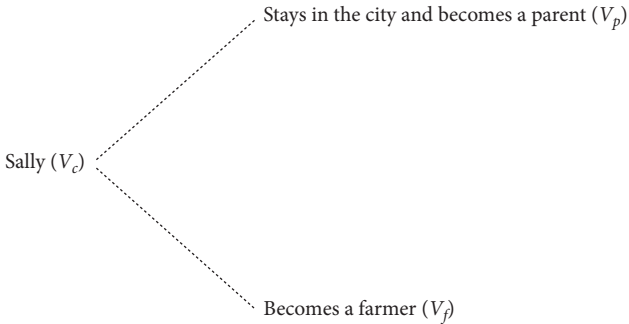


Figure 4.2 Sally's second choice

Actual-life evaluation tells us to evaluate both of Sally's futures in light of valuing attitudes that she would not have in *either* future. This is even more perverse. We should instead say that Sally's well-being as a parent is correlated with the values she would have as a parent, and that her well-being as a farmer is correlated with the values she would have as a farmer. In other words, we should adopt in-life evaluation.

4.5 Discovering What Is Good for You

We can now state the alienation test properly:

Alienation test If it is true that, were you to have X and to have no false beliefs about it, you would not desire it noninstrumentally, nor believe it to be good noninstrumentally, nor hope or wish to continue having it for its own sake, nor have a positive affective reaction to it, this is strong evidence that X would not be a constituent of your well-being. If it is true that, were you to have X and to have no false beliefs about it, you would desire it noninstrumentally, or believe it to be good noninstrumentally, or hope or wish to continue having it for its own sake, or have a positive affective reaction to it, this is weaker evidence that X would be a constituent of your well-being.

This is our best interpretation of the epistemic significance of Railton's idea. It is the result of the views we have taken about how to interpret 'valuing' and 'right circumstances'.

The alienation test underlines the importance of trying things out. So far as learning from your valuing attitudes goes, your best bet is often to try

things out and note your reaction to them. Otherwise, it is hard to draw any conclusions from the alienation test. Your future-directed current attitudes are an unreliable guide, since they do not take proper account of changing values.⁴⁴ For example, Sally's current attitudes towards living in the city or the country are not good evidence that living in the country would not be good for her.

This articulation of the alienation test helps to explain why prudence is difficult. Humans are able to predict the future to some degree and to anticipate their reactions to future events to some degree. Most obviously, we are often able to anticipate our reaction to events that are similar to those we have previously experienced, by using memories of those past experiences. Remarkably, we are also able to anticipate our reaction to completely novel events, simply by imagining them. For example, we know that 'chocolate pudding would taste better with cinnamon than dill', just by imagining it.⁴⁵ However, this 'prospective' ability is of course limited and subject to various errors.⁴⁶ For present purposes, the most important point is that our prospective judgements seem to be based on our current values:

People use their immediate hedonic reactions to simulations as predictors of the hedonic reactions they are likely to have when the events they are simulating actually come about. People do not imagine feeling anxious while having a colonoscopy so much as they imagine a colonoscopy, feel anxious, and then take this anxiety as an indicator of the feelings they can expect to experience during the procedure itself.⁴⁷

In other words, we imagine something and experience a reaction to the episode of imagination, and take this reaction as an indicator of the reaction we would have were the thing in question to come to pass. If this is how prospecting works, it is quite likely to lead us astray in cases involving changing values. When Sally imagines living in the country she may feel

⁴⁴ My proposal that we treat alienation as evidence about well-being resembles claims made in Hausman 2012: ch. 8. Hausman argues that current preferences are good evidence of what would be good for someone so long as the preferences are informed and self-interested. In my view, changing values make current preferences an unreliable guide even when they are well informed. In this respect, my proposal is closer to Sidgwick's formulation: '[a person's own good is] what would be desired... 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' Sidgwick 1907: 110–11.

⁴⁵ Gilbert and Wilson 2007: 1352.

⁴⁶ For discussion of these errors see Gilbert and Wilson 2007: 1354. See also Gilbert and Wilson 2000, Wilson 2002, and Loewenstein et al. 2003.

⁴⁷ Gilbert and Wilson 2007: 1352.

unhappy about the lack of social contact and amenities, and she may take this to be a guide to the feelings she would have were she to move there. If she does, she may well reach the wrong conclusion about whether moving there would suit her.

We can get only so far using our imaginations.⁴⁸ A person's best evidence of whether something would be good for her will often be obtainable only by trying it out.⁴⁹

4.6 Promoting Well-Being

Return now to the concern that motivated our discussion of how we can know what is good for us: how to promote well-being. We have some insight into things that, in general, benefit and harm people. We are pretty sure that health, education, autonomy, and good personal relationships generally make people better off. We are also pretty sure that ill-health, economic insecurity, and social isolation generally make people worse off. We have just concluded that often the best way for someone to find out what suits her individually is to try things out. What do these beliefs suggest about how can we promote well-being?

First, they suggest that we should try to ensure that people have easy access to the things that, generally, are good for people. It is reasonable to expect that people will do better when they are healthy, well-educated, autonomous, and have good personal relationships than when they are unhealthy, insecure, or socially isolated. This is not to say that they are bound to do better, or that people who lack these general goods must have bad lives. It is instead to say that, usually, people do better with these goods than without them. In this sense, we can reasonably expect that trying to ensure that people have these goods, or have easy access to them, will promote well-being.

Second, there are reasons to increase the ease, and decrease the cost, of trying things out. We can make prudence easier if we give people more

⁴⁸ Artworks are another possible source of imaginative insight into different circumstances. However, our reactions to artworks also seem likely to be based on our current values; if so, imaginative engagement with art, like prospection, seems prone to error in cases involving changing values. I am grateful to Tom Baker and Katharine Jenkins for discussion of this issue.

⁴⁹ '[S]tories, testimony, and theories aren't enough to teach you what it is like to have truly new experiences—you learn what it is like by actually having an experience of that type' Paul 2014: 13. Of course, trying things out is also not infallible evidence. As Daniel Haybron emphasizes, we often have imperfect understanding even of our current affective states. See Haybron 2008: ch. 10.

options and more security, since this would make it easier and less costly for them to discover what is good for them as individuals. The epistemic value of trying things out is, of course, to be set against the costs involved in trying things. Trying things is sometimes disruptive, sometimes disconcerting, sometimes irreversible, and always time-consuming. However, these costs are not fixed parameters of the human condition. We can do things to lessen them, such as providing and signalling opportunities to try things and providing security for those who try them so that they can back away from them if necessary. Insofar as we can decrease the costs of trying things out without undermining the epistemic value of doing so, there is a utilitarian case for doing this.

This generates an epistemic utilitarian argument for the importance of liberty and security. For utilitarians, the value of liberty goes beyond the value of being able to pursue one's own goals, of expanding the collective sense of possible modes of living, and of developing individuals' faculties.⁵⁰ It is also a matter of being able to discover what one values.⁵¹ Since discovery by trying things would often be too risky without security, these considerations also generate an epistemic utilitarian argument for security—where that means the capacity to cope if the thing in question turns out not to work well or not to suit the subject.

It does not follow that utilitarians should try to maximize the extent to which people try things out. Trying things is costly, and people have different preferences for novelty. The epistemic value of liberty and security needs to be considered in the context of the wider costs and benefits of choice and the ability to reverse past decisions.⁵² Merely having more choices, let alone exercising them, sometimes carries costs. Less obviously, there seem to be some downsides to reversibility of decisions.⁵³ Utilitarians should aim to optimize exploration and discovery, not to maximize it.

⁵⁰ 'Such are the differences among human beings in their sources of pleasure, their susceptibilities of pain, and the operation on them of different physical and moral agencies, that unless there is a corresponding diversity in their modes of life, they neither obtain their fair share of happiness, nor grow up to the mental, moral, and aesthetic stature of which their nature is capable' Mill 1991b/1859: 75–6.

⁵¹ In a similar vein, James Fishkin writes: 'Opportunities matter in part because they help each person formulate and revise his answer to the question of which paths and pursuits matter to him' 2014: 10, emphasis in original.

⁵² I am grateful to Zachary Hoskins and Aness Webster for discussion here.

⁵³ Gilbert and Wilson 2000: 188–93 point to psychological evidence that irreversible commitments sometimes trigger beneficial adaptation: what they call the 'psychological immune system'. The circumstances under which it is good to stop trying things out are modelled in mathematics as 'stopping rules': see Hill 2009.

Nevertheless, it might be that improving people's chances of knowing what is good for them is among the most important things we can do to promote utility. We can be led astray in thinking about utilitarianism by taking the idea of maximizing well-being too simply, as if it meant turning the well-being dial up as far as it will go. Of course, we do not have anything like that sort of control over well-being. A better metaphor may be searching for gold. We know some things about where to find it, but there may be no substitute for getting out there and digging.

4.7 Conclusion

It is disappointing that we do not yet know the nature of well-being. But it does not follow that we do not know anything at all about well-being. We know enough to form judgements about the plausibility of welfarism, and to reach significant conclusions about how to promote well-being.

Moreover, Railton's anti-alienation constraint can be used for epistemic rather than metaphysical purposes. We can use it to develop an account of an important way in which people can learn what is good for them, which I called the 'alienation test'. According to the best interpretation of this test, if you have something, and you have no relevant false beliefs about it, and yet it leaves you cold, that is strong evidence that it is not a constituent of your well-being; if under those circumstances you value it, that is weaker evidence that it is good for you. Because we have limited powers to anticipate our responses to circumstances that will change our values, this test implies that trying things is often an important way of discovering what we value, and what would be good for us. If this is correct, an important way to promote well-being is to make it easier and less costly for people to try things.

5

Two Kinds of Reasons

It is not much of an exaggeration to say that most criticisms of utilitarianism are one version or other of the idea that it is too simple. Our ethical convictions are complex and nuanced. If an ethical theory must match up to them fairly closely, it must have sufficient internal resources to generate this complexity and nuance. One common reaction to utilitarianism is that it does not have sufficient resources.

This reaction applies most forcefully to Act Utilitarianism, which is the simplest form of utilitarianism. One possible response is to abandon Act Utilitarianism in favour of some other version of Act Consequentialism. This is to seek extra complexity within the theory of value. But if we want more complexity than Act Utilitarianism offers we can seek it in more than one place, and we need to weigh up the merits of different ways of accommodating it. Since I find Act Utilitarianism's claims about value intuitively plausible, I will look in a different direction. Rather than explore alternative theories of value, I will explore alternative 'structures'—what remains of a theory of reasons or rightness when you remove its theory of value.

This chapter begins by considering Act Consequentialism, which is the structure of Act Utilitarianism. I will claim that Act Consequentialism goes wrong by failing to recognize an important kind of reason for action, which has to do with the contribution that each of us can make to good or bad patterns of action. I claim that this sort of reason exists alongside the sort of reason that Act Consequentialism recognizes, which has to do with the good or bad that each of us can do on each occasion that we act. The result is that I will advocate a pluralist theory of reasons, recognizing these two basic kinds. Since we have assumed that rightness is determined by reasons, this has implications for the theory of rightness, too. The resulting theory is somewhat similar to Rule Consequentialism, though with important differences that I will explain.

5.1 Act Consequentialism

If we are trying to reach reflective equilibrium in our ethical views, one of the things we are searching for is an ethical theory that consists of principles that are plausible in themselves. Derek Parfit wrote:

We have . . . intuitive beliefs, not only about which acts are wrong, but also about which principles or theories might be true. So as well as having plausible implications, any successful principle or theory must be in itself plausible. Only such a principle or theory could *support* our more particular moral beliefs.¹

As Parfit emphasizes, theorizing in ethics involves searching for principles or theories that are plausible quite apart from their implications about cases.

Act Utilitarianism does well in this respect. It is the result of three main claims, each of which is plausible in itself. These claims are Act Consequentialism, welfarism, and sum-ranking. Though it is more often presented as a theory of rightness, one way of understanding Act Consequentialism is as a claim about reasons for action.² Understood in this way, it claims that there is just one kind of reason for action: to make things go better. This is a very plausible claim when considered in the abstract. The whole point of acting, we might think, is to make some sort of causal intervention in the way things go—even if that means preventing something bad. If that is the point of acting, then it is plausible that the only genuine reason for action is to make things go better.³ As I noted in Chapter 1, welfarism and sum-ranking are also plausible claims. Welfarism claims that all well-being is noninstrumentally good, and that nothing else is. Though it is certainly possible to doubt welfarism, it is a plausible assumption. Similarly, sum-ranking is plausible. This is the claim that the value of an outcome is the sum of the goods (and bads) existing in that outcome.

¹ Parfit 2011: 415, emphasis in the original.

² It is presented as a theory of rightness by Smart 1973: 9, Sartorius 1975: 12, Brandt 1992: 9, Scheffler 1994: 1–2, Shaw 1999: 10, Hooker 2000: 5, and Portmore 2011: 34–8. In contrast, Norcross 2006 argues that the best form of Act Utilitarianism is a theory of the goodness of actions and outcomes. Some authors do present Act Consequentialism as a theory of reasons: see for example McElwee 2010b.

³ A. C. Ewing wrote: 'That it tends to produce a good seems not only a reason but the only ultimate reason for doing anything . . . This is at least a very plausible position to take up' 1948: 103. On the intuitive plausibility of Act Consequentialism see also Scheffler 1988: 1 and Foot 1985.

Each of the main components of Act Utilitarianism has some appeal independently of its implications about cases. However, many think that the implications of Act Utilitarianism about some particular cases are strongly counter-intuitive. We noted some criticisms of this sort in Chapter 2. For example, many think that Act Utilitarianism is wrong when it implies that one should break a promise whenever breaking it would generate slightly more well-being. Many also find it excessively demanding of agents.

Of course, some defend Act Utilitarianism from these claims.⁴ That is a worthwhile avenue for enquiry, leading to more sophisticated elaborations both of Act Utilitarianism and of the cases alleged to be problematic for it. Rather than go in that direction, however, I will explore other kinds of utilitarian theory. My hunch is that Act Utilitarianism lacks the resources to explain enough of our ethical convictions.

Act Utilitarianism consists of a simple theory of value, the conjunction of welfarism and sum-ranking, plus a simple structure, Act Consequentialism. If we think that it is too simple overall, which of these should we modify or reject?

Some philosophers are strongly disposed to hold on to Act Consequentialism, and seek to accommodate the complexity of our ethical convictions by adopting a more complex theory of value. Moreover, it may be that if we remove all restrictions from the theory of value, we can explain *any* set of judgements about rightness as arising from some version of Act Consequentialism.⁵ However, for some sets of judgements this would lead to a theory of value that is, taken by itself, hard to believe. For example, we may have to say that, so far as my reasons for action go, it is much worse for me to torture someone than for you to do so, and much worse for me to torture someone now than for me to torture someone later. No one would believe these claims about the agent-relative and time-relative badness of torturing if they were not needed—when taken together with Act Consequentialism—to explain the idea that there is a moral constraint against torture. Taken by themselves these claims about value are incredible.⁶

⁴ See Sartorius 1975: ch. 5 on Act Utilitarianism and promising. Kagan 1989 argues that Act Consequentialism is not excessively demanding.

⁵ See Louise 2004 and Portmore 2011: ch. 4 on the idea that all plausible moral theories can be 'consequentialized'. Schroeder 2007b, Brown 2011, and Schroeder 2017 doubt this claim.

⁶ Portmore 2011: 111–16 is a helpful discussion of, and reply to, this form of objection. As he makes clear (114), his advocacy of consequentializing is not a claim about what is logically possible, but a bet that some form of Act Consequentialism will turn out to be most plausible in reflective equilibrium. My bet is that a theory with a different structure will turn out to be more plausible.

Some convictions about cases can be modelled using a form of Act Consequentialism only by adopting a distorted theory of value, then. This is true in particular of convictions about moral constraints, which are hardest to reconcile with Act Consequentialism.⁷ Some Act Consequentialists think that, if this is true, it must cast doubt on those convictions. That could be the right response in some cases, but taken as a general policy it seems too sweeping. It amounts to being always more willing to amend one's theory of value or judgements about cases than to amend the structure of one's ethical theory. In contrast, when faced with implausible implications about cases I am inclined to modify Act Utilitarianism first by tinkering with its structure, and only later by considering modifications to its theory of value. In line with that strategy, in the rest of this chapter I will propose what I take to be a better structure for utilitarianism.

One clue about how to modify Act Consequentialism comes from an observation about the cases in which it seems problematic. Sometimes it seems to imply that we should break a promise or violate a right when, intuitively, these actions seem wrong. It also tells us that agents are required to sacrifice more when, intuitively, they are not required to do this. In other cases, Act Consequentialism endorses doing something that seems intuitively wrong because refraining from doing it would achieve little or nothing, due to compensating factors in the agent's environment. For example, we may think that it is wrong to make a journey by plane when it could be made quite easily in some more environmentally friendly way; but Act Consequentialism may imply that you should go by plane if someone else would take your seat in any case.⁸

In all of these cases, Act Consequentialism appears to go wrong in a specific way: by failing to account for a reason that we think exists. There appears to be a reason grounding a constraint or an option which Act Consequentialism fails to explain. The striking thing is that, though it could have gone wrong in the opposite way—by postulating a reason we think does not exist—it does not appear to do so. There *is* a reason to break the promise, to do more to help others, and to travel by plane. What is more, in each case Act Consequentialism seems to give the correct explanation of these reasons: there is a reason to do these things because doing them would

⁷ See Scheffler 1994: ch. 4 and Kagan 1989: chs. 1–4.

⁸ Markets often compensate for our choices in something like this way. There is also evidence that altruistic contributors to public goods reduce their contributions when others do more. See Fischbacher et al. 2001: 401. I am grateful to an anonymous reviewer of Woodard 2017 for this second point.

make things go better in some way.⁹ I conjecture that this is true in every case in which Act Consequentialism seems problematic.

In other words, Act Consequentialism appears to be *at least half right*. When it is possible to make things go better, there is always a reason to do so.¹⁰ This need not always be a decisive reason for action; perhaps sometimes we should not, all things considered, make things go better. But any alternative structure we adopt should imply that it is always a reason, nonetheless.

5.2 Pluralism

The reasons that Act Consequentialism recognizes have to do with the difference each action would make to the way the world turns out. If things would turn out better in some respect were you to go to Kalamazoo than were you not to go there, then you have a reason to go. If things would turn out best were you to go there, then according to Act Consequentialism you have most reason to go.

These are ‘act-based’ reasons. They are ‘act-based’ not in the sense that they are reasons to perform actions, but in the sense that the reason to perform the action is some property of the action itself. Since consequentialism claims that goodness is the sole reason-giving property, Act Consequentialism claims that the only kind of reason for action is that the action itself is good. The goodness of the action is the difference it makes (or would make, or would stand a chance of making) to the way things turn out.¹¹

The category of act-based reasons is extremely broad. Act Consequentialism is open-minded about what could make an action better or worse. It is compatible with the idea that actions can be good or bad in themselves, for example. If promise-keeping is good in itself, then there are act-based reasons to keep promises because of that goodness. This is what we should

⁹ See Kagan 1989: 15–19.

¹⁰ Note that this is a claim about Act Consequentialism, not a claim about Act Consequentialism when conjoined with a specific theory of value, such as hedonistic welfarism. Even so, it is of course controversial.

¹¹ The concept of act-based reasons is compatible with perspectival views of reasons: ‘goodness’ may mean ‘expected goodness’ or similar. It also allows that an action may be good because it has an objective probability between 0 and 1 of resulting in a good outcome. Finally, it is compatible with non-consequentialist theories of reasons, according to which the reason-giving property is something other than goodness. However, I will discuss act-based (and pattern-based) reasons in this book with only consequentialist theories in mind.

expect, given the open-mindedness of Act Consequentialism about value. Of course, we are interested in the idea as it could be applied within utilitarianism. This means that we are operating with a specific theory of value in mind, according to which the only thing that has noninstrumental value is well-being. The only act-based reasons the utilitarian can recognize are reasons to make people better off in terms of well-being. The category of utilitarian act-based reasons is thus significantly narrower than the category of act-based reasons.

We can state the idea of act-based reasons somewhat more precisely as follows:

ABR There is an act-based reason for agent *S* to do action *X* if and only if and because it is possible for *S* to do *X* and *S*'s doing *X* would be good.¹²

As this statement makes clear, there are two conditions for any agent to have an act-based reason to do something. That thing must be possible for her to do, and it must be good in some way. These conditions are individually necessary and jointly sufficient. Though it would be good for me to make the sun shine, this fact gives me no reason to do it, since it is not something I can do. But if I could make the sun shine, its goodness would give me a reason to do it.¹³

Act Consequentialism claims that all reasons are act-based in this sense. If we find Act Consequentialism plausible, it is partly because this claim about reasons is itself plausible. However, I think that there is another kind of reason, which has to do with another way we can think of actions themselves. Actions can be *parts* of other ethically significant entities—in particular, they can be parts of sequences or patterns of action. Some of these parthood relations, I claim, generate a different kind of reason.

To get a feel for what I mean, consider the following case. Suppose that you are all set to buy some gadget that is in short supply. You are at the front of the queue as the very last item in stock becomes available. Just as you are about to buy it, you remember that someone reliably informed you that this item was produced by children working in dangerous and exploitative

¹² Here and throughout, *X* may be an omission.

¹³ Here I assume that 'reason implies can'. For defence of this assumption, see Streumer 2007. Since I assume that no agent ought to do *X* unless she has a reason to do *X*, I also assume that 'ought implies can'. For doubts that ought implies can, see Sinnott-Armstrong 1984; for a reply to those doubts, see Streumer 2003. On how consequentialists should specify what it is possible for agents to do, see Bergström 1966, Carlson 1999, Bykvist 2002, and Portmore forthcoming.

conditions. You may decide not to buy it, for that reason. If so, you may have in mind the sort of reason I am pointing towards. It does not appear to be an act-based reason, since declining to buy would not be good in any significant way; it would have no effect on demand signals, for example, since the person behind you in the queue would immediately snap it up.

Still, you might say to yourself: 'I want no part in that.' You then seem to have in mind an example of what I will call a 'pattern-based' reason. Pattern-based reasons are due to a reason-giving feature of some possible pattern of action to which the agent could contribute; for consequentialists, they are due to the goodness of patterns. They reflect the significance of the fact that we can contribute to worthwhile activity through our actions in a way that is not reducible to the causal contribution our actions make. They are pattern-based in the sense that the goodness (say) of the pattern is the putative basis of the reason to perform the part. Like act-based reasons, they are reasons to perform actions: they are not to be confused with the reason some agent—a group agent, perhaps—may have to perform the pattern itself.

The two ways that an action may contribute to a pattern—causally, and as a part—are easy to conflate. Consider, for example, a couple's decision to have no more than two children out of concern for global overpopulation. A natural way to think of this decision, I claim, is as motivated by the couple's concern to play their part in a pattern of restraint that could be performed by potential parents across the world. Some such pattern would have good effects, let us suppose, by limiting the environmental damage caused by humans. The goodness of this broad pattern of restraint seems to generate some reason for this couple to limit the number of children they have. Of course, they also have an act-based reason to limit the number of children they have. Having two rather than three or four children itself reduces the burden on the environment. So far as the environment goes, let us assume, having fewer children makes things go better. More generally, there are often act-based reasons for participating in good patterns of action. This can make it harder to see that there are two kinds of reason.

However, we can see the difference between these two reasons to have fewer children by asking why the figure of two children is salient. This seems hard to explain in terms of act-based reasons. So far as the environment goes, there is just as much reason to have one child rather than two, three rather than four, or zero rather than one, as there is to have two rather than three. But some people seem to have the conviction that reasons for limiting the population generate, roughly, an option to have up to two children and

at least a weak constraint against having more than two.¹⁴ The obvious explanation of the salience of the number two that it is roughly the average number of children per couple that would keep the population stable. But if that is in the background, it suggests that people are thinking about a general pattern of procreative restraint, and of the goodness of this pattern as providing a reason to play their part in it.

In other cases the act of participation itself would achieve nothing, so the distinctness of act-based and pattern-based reasons is clearer. This is true in *compensation* cases, in which some feature of the agent's environment would neutralize the effects of her participation in a good pattern of action. The cases of refusing to buy a gadget and refusing to buy a plane ticket that we have already considered have this feature. If, as may be possible, refusing to buy the gadget or plane ticket would have no effect at all on the value of the relevant outcomes, we cannot appeal to act-based reasons to explain the apparent reason not to do these things. One possible explanation of these reasons uses the concept of pattern-based reasons instead.

We can state the idea of pattern-based reasons somewhat more precisely as follows:

PBR There is a pattern-based reason for agent *S* to do action *X* if and only if and because it is possible for *S* to do *X* and doing *X* is *S*'s part in an eligible pattern of action *P* that would be good.

This statement introduces the important concept of 'eligibility'. This is intended to mark the fact that we will want to recognize some constraints on which patterns generate reasons. Just as the fact that making the sun shine would be good does not suffice for me to have a reason to make the sun shine, the fact that the pattern of action <together, eliciting manna from heaven> would be good does not suffice for anyone to have a reason to play their part in that pattern—even if that part is something she could perform.¹⁵ Some patterns, though good, presumably do not generate any pattern-based reasons. For the moment we are simply calling these 'ineligible', though later we will consider the difficult question of what might make a pattern ineligible.

¹⁴ Mulgan 2006: 185–7 discusses this and similar numerical rules, arguing that they are not likely to be part of an ideal Rule Consequentialist code. However, the present point is that they enjoy some intuitive support amongst those concerned about population growth.

¹⁵ To make references to patterns easy to spot, I will place descriptions of them within angled brackets.

Granting that some patterns will be ineligible, we can afford to define the concept of a pattern of action in a very inclusive way. Thus let us say that any arbitrary combination of token actions is a pattern.¹⁶ Examples of patterns then include the following:

- <all potential parents exercise procreative restraint>
- <all road users in the UK drive on the left-hand side of the road>
- <all promise-makers keep their promises>

In these examples, the patterns are both regular in the sense that each participant does the same kind of thing, and salient in the sense that the patterns look like the kind of thing that we may be interested in when thinking about ethics. But I said that any combination of actions is a pattern, so we should note some irregular and non-salient examples, too:

- <I mow the lawn, you eat a gherkin>
- <Henry VIII scratches his nose, the first child born in the year 2150 learns to play the trumpet>

As these examples illustrate, a ‘pattern of action’ in the sense that I intend can be any arbitrary combination of token actions, including ones performed in the past or the future. The purpose of throwing the doors open in this way is to avoid prejudice in considering the scope of the idea of pattern-based reasons. We delimit the scope self-consciously by considering eligibility, later.

Now, as my statements of the ideas may have made obvious, we can think of act-based reasons as a special case of pattern-based reasons. They are the special case that results when the pattern *P* is identical to *S*’s part, *X*. *S* then has a reason to do *X* which is due to its being *S*’s part of the good pattern *X*—assuming that identity is a limiting case of parthood. In this sense, the idea of pattern-based reasons is a more general idea than the idea of act-based reasons.¹⁷ This does not, of course, do anything to show that there are any pattern-based reasons other than act-based reasons; for all it shows, perhaps there are none. But it is a notable theoretical connection between the concepts, and later I will appeal to it when discussing the eligibility of patterns.

There are at least four arguments for believing that there are pattern-based reasons (other than act-based reasons: I will drop that qualification

¹⁶ Compare Jackson 1987: 93.

¹⁷ Compare Bacharach 1999: 118.

henceforth). The first is that the idea appears to be part of our ordinary ethical beliefs, and that this gives us *prima facie* reason for thinking that it is correct. In ordinary life, people often express the idea that there is a reason to contribute to good patterns of action, and a reason not to contribute to bad patterns of action. Admittedly, they could be pointing only to act-based reasons to contribute or not to contribute. But this is not how things seem. They say things like ‘I will play my part’, ‘I want no part in that’, ‘you should do your bit’, and ‘you should not be complicit’. These expressions draw attention to the relationship between the thing to be done or not done and some larger pattern of which it is part. A natural interpretation of these expressions is that they are raising the salience of that relationship because the speakers assume that it will be taken to be part of the reason for or against the action under discussion. The idea of pattern-based reasons appears to be prevalent in ordinary ethical thought and discussion.¹⁸

Second, the idea seems plausible when stated in the abstract. Admittedly, I can do no more here than report what seems plausible to me. But the idea that one kind of reason for action is to contribute to a larger pattern of worthwhile activity seems, to me, roughly as plausible in the abstract as the idea that another kind of reason for action is to make things go better.¹⁹ The second idea may be more familiar than the first, because it has been discussed so much by philosophers. But the idea that one reason for doing something is to join in with others in doing something worthwhile, or that one reason for *not* doing something is to avoid joining in with others in doing something bad, also makes sense. It reflects a way of thinking of actions that is not alien or contrived: they can fit together with others to form larger patterns, and their being part of these larger entities can be an ethically significant fact about them of the sort that could generate a reason. For example, the fact that having no more than two children is my part in a good pattern of procreative restraint by parents and potential parents may be a consideration in favour of my having no more than two children.

Third, the idea keeps being reinvented by philosophers and economists thinking about reasons for action. Arguably it is present in Kant’s idea that each of us should act as if a member of the Kingdom of Ends: we should each

¹⁸ Admittedly, ordinary expressions such as ‘I must do my bit’ can be interpreted in different ways, including ways that make reference to non-consequentialist ideas. I take this first argument to be suggestive, but by no means conclusive.

¹⁹ Note that I am here using ‘pattern-based reasons’ to refer to pattern-based reasons other than act-based reasons. On this narrow usage, act-based reasons are not a special case of pattern-based reason.

play our part in the possible pattern of action <every rational agent never fails to treat any rational agent with respect>.²⁰ It certainly seems present in Collective Consequentialism and, as I shall discuss shortly, in certain kinds of Rule Consequentialism.²¹ Shifting intellectual contexts, it is one of the constituent ideas of Team Reasoning—which has seemed to some economists and decision theorists to be the right way to make sense of how people should solve coordination problems.²² There is also evidence from behavioural economics that people in fact employ Team Reasoning when faced with coordination problems.²³ Finally, in another quite different context, the idea of pattern-based reasons is a constituent of the doctrine of possibilism in deontic logic, and of Edward McClennen’s associated idea of resolute choice in decision theory.²⁴ It is to the idea’s credit that it has seemed to make sense to thinkers—often working independently of each other—in these quite different areas.

The fourth argument for belief in pattern-based reasons is that it offers to provide the best explanation of moral constraints. The problem with constraints—and the reason why they are often said to be paradoxical—is that it seems impossible to explain them by pointing to the badness of the activity they constrain, yet it seems as though the badness of this activity is precisely the thing that ought to explain the constraint. For example, it seems impossible to explain the constraint against torturing people by pointing to the badness of torture: if torture is so bad, it should be minimized; this means that we should be willing to torture if that is necessary to minimize it; but then there is no constraint against torturing, only a reason to minimize it.²⁵ Now, we can try to explain constraints in other ways. We

²⁰ See Woodard 2013. Kantians sometimes use the idiom of ‘playing one’s part’ to explain Kant’s ideas. For example, in criticizing Kant’s claim that one should not lie even to a murderer, Christine Korsgaard says: ‘in cases such as that of the murderer at the door it seems grotesque simply to say that I have done my part by telling the truth and the bad results are not my responsibility’ 1996: 150. I am grateful to Guy Fletcher for pointing out the relevance of this passage.

²¹ On Collective Consequentialism see Mulgan 2001. On Rule Consequentialism see Section 5.4.

²² Team Reasoning adopts both the idea of pattern-based reasons and the idea that patterns should be evaluated from the perspective of the team. See Sugden 1993 and Bacharach 1999 and 2006.

²³ See Colman et al. 2008 and Bardsley et al. 2010.

²⁴ Possibilists claim that what an agent ought to do depends on how she *could best* respond to her own actions, rather than (as actualists claim) how she *would* respond to them. See Jackson and Pargetter 1986, Zimmerman 1996: ch. 6, and Woodard 2009. Resolute choice involves a kind of intrapersonal cooperation over time: see McClennen 1990 and Woodard 2008a: 54–8.

²⁵ See Scheffler 1994: ch. 4 and Kagan 1989: chs. 1–4. Two attempts to show that constraints are not paradoxical are McMahan 1991 and Heuer 2011.

could simply postulate a moral principle forbidding torture, for example, or we could appeal to the idea of agent-relative and time-relative value. But intuitively these explanations cite the wrong things. The most important part of what is terrible about torture is what it does to its victims, which is a matter of agent-neutral value.

The idea of pattern-based reasons seems to make it possible to explain this and other constraints by appeal to agent-neutral value. The central difficulty in explaining constraints is finding some way of singling out the agent's present action. As we noted, attempts to do this by attaching extra (agent-relative and time-relative) value to it just distort our claims about value. But the idea of pattern-based reasons already singles out the agent's present action, since it refers to her part (now) in the larger pattern. In the case of torture, for example, we can hypothesize that the agent is under a constraint not to torture now because this is her part, now, in the general pattern of action <no one tortures anyone>. The goodness of this pattern is entirely agent-neutral, as it should be; it is the parthood relation, not any claim about value, that singles out the agent's present action.²⁶ Similarly, my pattern-based reason not to make a short-haul journey by plane is that this is my part in a good general pattern of reducing emissions: the goodness of this pattern is agent-neutral, but the pattern-based reason singles out my action rather than supporting a reason to minimize use of air travel. We get the desired explanation via an independently plausible structural feature of our ethical theory rather than by distorting our theory of value.

These four arguments fall a long way short of proof that pattern-based reasons exist. They each raise further controversial issues and invite possible objections or doubts. What I hope they do is to motivate interest in the idea of pattern-based reasons. In later chapters I will make repeated use of the idea, and it may be that the use to which it is later put will provide some additional reason for believing it—or, just as welcome, will suggest to others some better ways of employing it.

If we do come to believe in pattern-based reasons, however, it will not make sense to believe in them alone. We should still believe in act-based reasons. If the parthood relation between an action and something good can provide a reason to perform that action, it is hard to deny that the causal relationship between an action and something good can provide a reason to perform it. As we concluded earlier, Act Consequentialism is at least half

²⁶ Of course, there are likely to be strong act-based reasons not to torture, as well.

right—in the sense that all of the reasons that it claims exist really do exist. If we believe in pattern-based reasons, then, we should be pluralists about kinds of reason.²⁷ In the remainder of this book we will put this idea to work, exploring some of the issues that arise when we try to develop a pluralist account of reasons. The first of these is the issue of the eligibility of patterns.

5.3 The Minimal Constraint on Eligibility

Recall that we defined a pattern of action as any arbitrary combination of token actions. We then introduced the concept of eligibility to mark the distinction between those patterns that generate pattern-based reasons and those that do not. Hence one of the challenges facing any theory of pattern-based reasons, including pluralist theories, is to give some account of what makes patterns eligible.

A minimal constraint on eligibility, I will assume, is that the whole pattern could be performed. That is, it must be true of each of its parts—the constituent actions—that it could be performed if the other parts were performed. If a pattern is not performable in this sense it is very hard to see how its goodness could give anyone a reason to play their part in it—though, of course, they may have other reasons for playing their part in it. This constraint on eligibility is the analogue of the claim that there can be no act-based reason for *S* to do *X* if *S* could not do *X*. In fact, this condition on the existence of act-based reasons follows from the requirement on pattern-based reasons that the pattern be performable, in the case where the agent's part is identical to the pattern.

So far, the accounts of act-based and pattern-based reasons are uniform. The major question is whether there are additional constraints on the eligibility of patterns. Now, it may seem obvious that there must be additional constraints. To see this, consider what we might say about reasons to keep promises in the following case.

Whenever John makes a promise, his friend Jim threatens to reveal a somewhat embarrassing secret about him if he keeps the promise. Jim carefully calibrates the threat so that it is just bad enough to make it all things considered better, impartially speaking, for John to break each promise than to keep it. John's past contains enough embarrassing secrets, of

²⁷ Note that this is quite different from perspectival/non-perspectival pluralism, which we considered in Chapter 3.

sufficient variety, to sustain this practice indefinitely. Suppose that we think that, nevertheless, John should keep his promises, and we are trying to use the concept of pattern-based reasons to explain why this is so. We hypothesize that the value of keeping promises, as a general pattern of action, may give John an additional reason to keep his promise, since that is his part in this pattern. In other words, we cite the following pattern, or something similar, to explain an extra reason for John to keep his promise:

<Everyone keeps his or her sincerely made promises, when possible>

This pattern is consistent with the minimal constraint, since every part of it could be performed. But many other, apparently irrelevant, patterns are also consistent with the same constraint:

<I mow the lawn, you eat a gherkin, John keeps his promise>

<Henry VIII scratches his nose, the first child born in the year 2150 learns to play the trumpet, John keeps his promise>

We do not want to say that John has a reason to keep his promise associated with each of these patterns, though each has some value. So, it may seem, we must accept additional constraints on eligibility.

In fact, it may be possible to exclude these patterns without adopting any additional constraints. Once again it is helpful to draw an analogy with the way we think of act-based reasons. When thinking about those, we discount irrelevant elaborations and extensions of options. Suppose that you confront a trolley problem, and we say that you have an act-based reason to *pull the lever to switch the trolley*, since that will save four lives. We do not worry about whether you have a reason to *pull the lever to switch the trolley while whistling*, though that would also save four lives. Nor do we worry about whether you have a reason to *pull the lever to switch the trolley and later go to the supermarket*, though that too would save four lives. We discount the first because it is an irrelevant elaboration of your option, and the second because it is an irrelevant extension of it.²⁸

²⁸ Portmore forthcoming: ch. 4 advocates ‘maximalism’, the view that what an agent ought to do is a function, not of all her options, but of her ‘maximal options’—that is, maximally specific options—in particular. This may sound at odds with my claim that we discount irrelevant elaborations or extensions of options. However, Portmore specifies maximalism in terms of ‘maximally evaluatively specific options’. In our example, *pulling the lever* is a maximally evaluatively specific option in Portmore’s sense, since the only actions that entail it (such as *pulling the lever while whistling*) are evaluatively identical to it.

It is not entirely clear how to characterize the exclusion of irrelevant elaborations and extensions. On one possible view, they are excluded by an additional constraint on eligibility of some sort, so that they are not eligible options, strictly speaking. On another, they are not excluded by any constraint, but are simply ignored in speech and thought about options, since they are taken not to be of any independent interest. Whichever is the correct description in the case of act-based reasons, though, we can presumably treat pattern-based reasons in the same way. Intuitively, the pattern involving Henry VIII and the pattern involving your eating a gherkin are both irrelevant extensions of the minimal pattern <John keeps his promise>. He has reasons associated with that minimal pattern—his act-based reasons—and, if these other patterns are irrelevant elaborations or extensions of it, we can simply discount them.

We can exclude irrelevant extensions and elaborations, then. But the central question is whether this excludes enough. In particular, some believers in pattern-based reasons claim that a further constraint on eligibility is that the pattern in question stands some chance of being realized if *S* were to play her part. On this view, if the other agents needed to realize the beneficial pattern would not play their parts, the pattern is ineligible. This is to adopt what we might call a ‘willingness requirement’ on the eligibility of patterns.

One version of this requirement is a feature of Donald Regan’s Cooperative Utilitarianism. According to this theory, each agent ought to ‘*co-operate, with whoever else is co-operating, in the production of the best consequences possible given the behaviour of non-co-operators*’.²⁹ Now, this is a claim about oughts, not reasons. But we can borrow Regan’s idea, and consider it as a possible constraint on the eligibility of patterns. We can state a schematic version of this possible constraint as follows:

Willingness There is a pattern-based reason for agent *S* to do action *X* because it is his part in some valuable pattern of action *P* only if the other agents required to realize *P* would cooperate in doing so, were *S* to do *X*.

This is merely schematic, because we do not yet know what is required in order to ‘cooperate’.³⁰ On a purely behavioural conception, cooperating requires only that each agent perform the behaviour necessary for *P* to be

²⁹ Regan 1980: 124, emphasis in the original. Other endorsements of versions of the willingness requirement include Hurley 1989: 146, Bacharach 2006: 127, and Dietz 2016: 974. In Section 5.5 I will endorse a version of it.

³⁰ See Regan 1980: chs. 8–10. Regan argues that being cooperative requires using a certain decision procedure, not merely exhibiting the requisite behaviour.

realized. On more demanding conceptions, cooperating requires that each agent have, in addition, certain intentional states—such as, for example, the intention to act jointly with others in producing *P*.³¹

Thus there are several importantly different versions of the willingness requirement. It is also important to emphasize that versions of the willingness requirement are not the only interesting candidate constraints on eligibility that go beyond the minimal constraint. Nevertheless, a major question for any theory of pattern-based reasons is whether any version of the willingness requirement is a constraint on the eligibility of patterns. The answer to this question determines whether the theory conceives of pattern-based reasons as arising only in cooperative contexts, in some sense of ‘cooperative’.

5.4 Rule Consequentialism

Some idealizing forms of Rule Consequentialism can be interpreted as theories of pattern-based reasons. If they are correctly so interpreted, these versions of Rule Consequentialism are the most prominent theories of pattern-based reasons that do not adopt any version of the willingness requirement. For this reason, it is helpful to consider some of their features here. In Sections 5.5 and 5.6 I will explain how my own view differs from idealizing Rule Consequentialism.

By an ‘idealizing’ form of Rule Consequentialism, I mean one that defines the value of a set of rules using some idealizing assumption, such as that the vast majority of people comply with the rules.³² This assumption is idealizing in the sense that it is not supposed to be realistic; it is incorporated in the theory for some reason other than the supposition that it accurately describes the world. When Rule Consequentialists of this sort evaluate sets of rules on the assumption that almost everyone complies with them, they are not being naive. Instead, they are being deliberately idealizing. They are claiming that it is morally appropriate to evaluate rules in this way.

³¹ See Woodard 2017: 111–15. Different analyses of joint action yield different specifications of the requisite intentional states. Another variable in specifications of the willingness requirement concerns the probability of cooperation. I will ignore these complications here.

³² There has been much recent discussion about the most plausible form of idealization. See Hooker 2000: 80–5, Mulgan 2001: ch. 3, Mulgan 2006: ch. 5, Ridge 2006, Smith 2010a, Tobia 2013, Toppinen 2016, Mulgan 2017, Yeo 2017, and Podgorski 2018.

The question, then, is why this might be morally appropriate. One way for Rule Consequentialists to answer this question is to construe their theories as theories of pattern-based reasons whilst rejecting all forms of the willingness requirement. They can say, for example, that in evaluating a rule such as ‘never lie’ on the assumption that almost everyone complies with it, they are asking whether the pattern of action consisting of almost no-one lying is a good one, and they are treating it as eligible to generate pattern-based reasons. If the rule ‘never lie’ is a member of the best set of rules, then each agent has, according to these forms of Rule Consequentialism, a reason to play her part in this good pattern, by not lying herself. This would be to say that each agent has a reason to play her part in the best that (nearly) all agents could do, with respect to lying—even if others are not playing their parts.³³

This interpretation of idealizing Rule Consequentialism illuminates a number of its features. For example, leading versions of Rule Consequentialism adopt a ‘disaster avoidance rule’, according to which it is right to break other rules in the best set when necessary to avoid a disaster.³⁴ In Brad Hooker’s example, the disaster avoidance rule tells me to break my promise to return my neighbour’s chainsaw to him if I can see that he has murderous intentions.³⁵ What is my reason for breaking this promise? Presumably, it is that breaking it is very likely necessary to avoid terrible consequences on this specific occasion: it is an act-based reason.³⁶ If so, these forms of Rule Consequentialism are, in fact, pluralist theories of reasons, of the kind we are exploring in this chapter. They combine pluralism with additional claims about the relative strength of reasons, to the effect that, when pattern-based reasons conflict with act-based reasons, the pattern-based reasons are usually stronger, but can be defeated when a disaster is on the cards.

Second, interpreting Rule Consequentialism as a theory of pattern-based reasons helps to answer the objection that it is either incoherent or

³³ See Woodard 2008b. I do not claim that all versions of Rule Consequentialism are properly interpreted as theories of pattern-based reasons. In particular, versions that define the value of a set of rules in terms of something other than compliance with them—such as those versions that define it in terms of acceptance of the rules, including Hooker 2000—may not amount to theories of pattern-based reasons. This depends on whether it is illuminating to think of their claims about reasons in terms of patterns of similarly motivated actions.

³⁴ For example, see Brandt 1992: 151 and Hooker 2000: 98–9.

³⁵ Hooker 2000: 98–9. Note that Hooker evaluates rules in terms of the value of acceptance of them, not of compliance with them (75–80).

³⁶ In fact, this is not the only act-based reason recognized by plausible forms of Rule Consequentialism. They also recognize a general reason to benefit people, though this reason will often be subordinate to others when it conflicts with them. See Hooker 2000: 98 n. 7.

extensionally equivalent to Act Consequentialism. This objection takes the form of a dilemma. Either Rule Consequentialism is extensionally equivalent to Act Consequentialism, or it is not. If it is, it cannot do a better job of matching and explaining our convictions than does Act Consequentialism, and for practical purposes it is not a distinct theory. On the other hand, if it is not extensionally equivalent to Act Consequentialism it must imply that it is sometimes right to fail to maximize the good. If so, it cannot be a coherent form of consequentialism.³⁷

For the moment, assume that Rule Consequentialism is not extensionally equivalent to Act Consequentialism. Under this assumption the accusation is that Rule Consequentialism is incoherent, since it implies that it is sometimes right to fail to maximize the good. This is an extremely strong claim. It is one thing for a moral theory to be false, or wrong-headed; quite another for it to be, strictly, incoherent. Yet this remains a common claim about Rule Consequentialism.³⁸

How exactly is incoherence supposed to arise? The most straightforward way would be to assume that *consequentialism* itself is extensionally equivalent to Act Consequentialism, at least in its assessment of actions. Then any form of consequentialism that fails to be extensionally equivalent to Act Consequentialism must be incoherent.³⁹ But, though some who make this objection may indeed be relying on this assumption, it does not make for an interesting objection. Without any reason for thinking that all forms of consequentialism must be extensionally equivalent to Act Consequentialism, it amounts at best to definitional fiat.⁴⁰ More charitably, we may suppose that those who level this objection just cannot see why it would make sense for a consequentialist to claim that an act's rightness depends on the value of something *other* than the act itself. That is, perhaps they cannot

³⁷ See Smart 1973: 9–12 and Lyons 1965: 142. My discussion of this objection is indebted to Hooker's: see Hooker 2000: ch. 4. My explanation of how Rule Consequentialists can answer the objection is different than his, however.

³⁸ For example, it is discussed in Law 1999 and Wolf 2016.

³⁹ This appears to be why Susan Wolf takes the objection seriously. She writes: 'If, as utilitarians think, the moral point of view is the point of view that assesses everything according to *its* tendency to bring about the greatest good for the greatest number, then it will yield both the judgment that the best (and so the morally right) rules will be the rules that maximize utility and the judgment that the best (and so the morally right) actions will be the actions that maximize utility' Wolf 2016: 132, my emphasis.

⁴⁰ As Frances Howard-Snyder puts it: 'Of course, if the rule utilitarian accepted the act utilitarian view that an act is right if and only if it maximizes utility, then she would be inconsistent when she told agents to do things that didn't maximize utility. But perhaps it is unfair to weigh her down with a view that will so quickly cause her to sink' 1993: 275.

see any intelligible rationale for indirect consequentialist evaluation.⁴¹ Finding indirect evaluation impossible to rationalize, they use a principle of charity to infer that Rule Consequentialists must really believe in the direct evaluation of actions, and thus conclude that they are incoherent after all.

To answer this objection we need to find a rationale for indirect evaluation of acts in terms of the value of rules. If we interpret Rule Consequentialism as a theory of pattern-based reasons, we can provide one. It makes sense to evaluate actions in terms of the value of rules because, in addition to being causes, actions are parts of patterns of action. If there are pattern-based reasons in virtue of these parthood relations, we must ask which patterns generate them—that is, which are eligible. Some or all of the eligible patterns may consist of behaviour that complies with a rule or set of rules. If so, the value of these rules will be relevant to the strength of the pattern-based reasons they generate. So, if we wish to take account of all of the reasons there are, one thing we should do is to ask what the value of these rule-specified patterns is—which is the same thing as asking what the value of the rules would be, were they complied with.

The accusation of incoherence is clearly misplaced. There are certainly difficulties in developing a theory of pattern-based reasons, as I shall discuss shortly. Even if we were to develop a full-fledged theory of them, the theory may be wrong. It may be that actions do not have the ethical significance attributed to them by the concept of pattern-based reasons. But it is simply false to say that there is no possible rationale for indirect consequentialist evaluation of actions in terms of the value of rules. Thus the allegation that Rule Consequentialism is committed to only direct evaluation of actions, and so is ultimately incoherent, is simply mistaken.⁴²

Return then to the question of whether Rule Consequentialism is extensionally equivalent to Act Consequentialism. The ‘collapse’ into extensional equivalence could happen in two ways. First, it could be that the best set of rules instructs agents to perform the action with the best consequences, under that very description (either because the only rule it contains is Act Consequentialism itself, or because the rules it contains have exception clauses which tell agents to perform the action with the best consequences). Extensional equivalence would then follow from intensional equivalence. However, for compliance versions of Rule Consequentialism this will not happen, because Act Consequentialism fares less well in solving

⁴¹ See Kagan 1989: 37.

⁴² Hooker 2000: 99–101 is a more sympathetic discussion of the objection.

coordination problems than theories directing the agent to play her part in the best that she could do along with other cooperative agents. In coordination problems there is more than one equilibrium—that is, more than one outcome in which it would be true that each agent has performed the ‘best reply’ to the actions of everyone else. So long as an equilibrium is reached, each person has made the outcome as good as she could make it. But some equilibria are better than others, and pattern-based reasons would direct agents towards these superior outcomes.⁴³

Second, extensional equivalence would occur if the best set of rules were to instruct agents to perform the action with the best consequences despite not picking out this action under that description. This might be true if, for example, every rule contained an exception clause for each case in which the rule recommends something sub-optimal, and the exception clauses all pick out those actions that would be best (albeit under some other description). For example, suppose that the best set of rules contains a rule governing promise-keeping. The thought would be that, in order to be the best rule governing promise-keeping, it would have to make exceptions for cases in which breaking a promise would have better consequences than keeping it. If a rule recommends doing the thing that would have best consequences in every case, it will be extensionally equivalent to Act Consequentialism even though it is not intensionally equivalent.

On our current assumption, however, the exception clauses must pick out features of the exceptional cases under some description other than ‘what would be best in these circumstances’. They might, for example, say things like ‘do not keep your promise if keeping it would lead to loss of life’, and so on for other kinds of exception. These clauses might distinguish very many kinds of exception, and specify what the agent ought to do in terms of different features for each kind of exception. But however fine-grained these exception clauses are, which actions they recommend will depend on the circumstances used to evaluate the set of rules. For idealizing versions of Rule Consequentialism, these will be idealized circumstances of one sort or another. But there is no reason to think that an exception clause that

⁴³ Gibbard 1965 pointed out that even if all agents in a coordination game comply with Act Utilitarianism, a sub-optimal outcome may result. Regan 1980 developed the point and showed that his theory Cooperative Utilitarianism handles coordination games better. Note that, under circumstances of *universal* compliance, the simple rule ‘play your part in the best that everyone could do’ does just as well as Regan’s more complex theory. For acceptance versions of Rule Consequentialism there are further reasons why the best set of rules would not contain only Act Consequentialism. See Hooker 2000: 94, 142–4.

requires the agent to do what is best in ideal circumstances, will also require the agent to do what is best in non-ideal circumstances.⁴⁴ There is no reason, that is, unless the rule in question instructs the agent to ‘do what is best’ under that very description.

So we should not expect the best set of rules to be extensionally equivalent to Act Consequentialism. This collapse does not occur due to intensional equivalence, because ‘do what is best’ fares worse than alternative rules, such as Cooperative Utilitarianism, in coordination problems. It does not occur without intensional equivalence, because exception clauses that are extensionally equivalent to Act Consequentialism in ideal circumstances will not, in general, be extensionally equivalent to it in non-ideal circumstances.

However, note that this reasoning depends on the assumption that rules are evaluated according to some idealizing assumption. In large part, it is the fact that standard forms of Rule Consequentialism are idealizing that explains their extensional non-equivalence to Act Consequentialism. We will return to this point in the next section.

The real question for Rule Consequentialists, then, is not whether their theory is incoherent, or whether it is a genuine alternative to Act Consequentialism. Those old objections are misplaced attempts to prevent Rule Consequentialism from getting out of the starting blocks. The real question is whether Rule Consequentialism is, all things considered, the most defensible theory. And one part of the answer to that question depends on whether Rule Consequentialism’s idealizing stance on the eligibility of patterns can be defended. In the remainder of this chapter I shall argue that it cannot.

5.5 Accepting the Willingness Requirement

In previous work I have argued against all forms of the willingness requirement. I claimed that rejecting the willingness requirement enables us to give a promising explanation of moral constraints, and that the usual reasons for accepting it—that it helps to explain why mere agglomerations are not eligible patterns, and that it is necessary to explain why we should not take futile or dangerous actions—are not convincing.⁴⁵ The result was that I accepted a pluralist view quite close in character to idealizing Rule Consequentialism.

⁴⁴ See Woodard 2008b: 243 n. 5 and Podgorski 2018: 281.

⁴⁵ See especially Woodard 2008a and Woodard 2017.

However, a recent paper by Alexander Dietz has changed my mind about the willingness requirement.⁴⁶ Because of Dietz's arguments, I now believe that a weak form of the willingness requirement *is* a constraint on the eligibility of patterns. According to this requirement, a pattern is ineligible unless there is at least some chance that it would be realized. Happily, I now also believe that it remains possible to offer a plausible explanation of moral constraints even while accepting this constraint on eligibility.

Dietz's arguments focus on what a theory of pattern-based reasons that rejects the willingness requirement could say about the *strength* of reasons. He compares Bernard Williams's case of Jim with a case in which two agents could paint a house together.⁴⁷ Jim faces a choice between shooting one innocent person himself, or not doing so—in which case Pedro will shoot that person plus nineteen more. In the house-painting case, two people could paint the house, and it would be good for them to do so, but one of them is unwilling.

Now, a pluralist theory of pattern-based reasons that rejects the willingness requirement may seem to offer a promising explanation of our convictions about Jim. It can say that Jim has a strong act-based reason to shoot, since doing so would save nineteen lives. But it can add that he also has a pattern-based reason not to shoot, since it is possible for Jim not to shoot and for Pedro not to shoot, and this is the best that both could do, and not shooting is Jim's part in this valuable pattern of action. This is in some ways an attractive explanation of the conviction that Jim is subject to conflicting reasons. Instead of locating the source of Jim's reason not to shoot in *Jim's* integrity, or in his special responsibility for what he does rather than allows, it locates it in *Pedro's* agency. It is because the nineteen are imperilled by another agent and not by a non-agential natural force that Jim has a reason not to shoot.⁴⁸

But if we say that Jim has a reason to play his part in the pattern <neither Jim nor Pedro shoots anyone> despite Pedro's unwillingness to cooperate, it seems as though we should also say that the willing agent has a reason to play his part in the pattern <willing agent and one other paint the house together> despite the unwillingness of the other. That might not be too bad, if this is a trifling reason that no one should take into account in

⁴⁶ See Dietz 2016: 969–72.

⁴⁷ Dietz 2016: 970–1. The case of Jim is described in Williams 1973: 98–9.

⁴⁸ See Woodard 2008a: ch. 2. Frances Kamm has distinguished between agent-focused and victim-focused explanations of moral constraints: see Kamm 2007: 28–9. The claim that Jim's reason not to shoot is a pattern-based reason is a third kind of explanation, which is focused on the contrast between (a) other agents and (b) other features of the agent's environment. The significance of this contrast is also emphasized in Alexander 1985.

deliberation, let alone act upon. But here we reach the crux. What could explain why Jim's reason not to shoot is strong enough to account for our conviction that he faces a difficult moral decision, while the willing painter's reason is trifling? It seems as though it must have something to do with the values of the relevant patterns: saving twenty lives is very much more important than painting the house. So, it seems as though anyone holding a view like this should accept that the strength of pattern-based reasons increases with the value of the pattern concerned, other things equal.⁴⁹

Once we add this hypothesis about the strength of reasons we can derive an unacceptable implication from the resulting view. For example, suppose that war and disease would be eradicated permanently if Wendy and Hans were both to press a certain button in the next five minutes. If only one presses, an innocent person will be killed and war and disease will continue unchecked, while nothing will happen if neither presses the button. Suppose that Wendy is willing to press but that Hans is not. If we reject the willingness requirement but we accept that the strength of reasons increases as the value of the relevant pattern increases, we have to accept that at some point Wendy's reason to press, despite Hans's unwillingness, will be stronger than her reason not to press. This seems to imply that she ought to press the button. But given that the certain outcome of Wendy alone pressing the button is the futile loss of one innocent life, it is very hard to believe that she ought to press it.⁵⁰

I agree with Dietz that this is an unacceptable implication. It is the result of combining a theory of pattern-based reasons that rejects all forms of the willingness requirement, on one hand, with the hypothesis that the strength of pattern-based reasons increases as the value of the relevant pattern increases, on the other hand. One option, therefore, would be to reject this hypothesis about the strength of reasons. That would enable us to restore consistency between the theory and the conviction that, in the problematic case, Wendy ought not to press the button unilaterally.

Perhaps some version of that response can be made to work. However, I am not inclined to pursue that route. What we want is not mere consistency with the relevant convictions, but an explanation of them. So far as I can see, the hypothesis about the strength of pattern-based reasons upon

⁴⁹ For Dietz's argument to work we do not need to assume, as he does (2016: 971), that the strength of pattern-based reasons varies in a fixed proportion to the value of the relevant pattern. However, we do need to assume that it does not approach an upper bound asymptotically. This second assumption is not trivial, but it seems to me to be correct.

⁵⁰ See Dietz 2016: 971–2. One wrinkle is that utilitarians need not assume that agents ought always to perform the action for which there is strongest reason. I discuss this briefly in Chapter 9, this volume. However, we can ignore this complication here.

which Dietz relies is very likely to be endorsed by any acceptable explanation of the strength of reasons. In the absence of any plausible alternative account of the strength of reasons, I think that we should accept the hypothesis.

If all of that is correct, then theories of pattern-based reasons should endorse some form of willingness requirement, and so accept that there are no pattern-based reasons in uncooperative contexts. This moves them away from idealizing forms of Rule Consequentialism. In doing that, it makes them less susceptible to worries about the futility of complying with rules when others are not. But as we saw in Section 5.4, the fact that standard forms of Rule Consequentialism are idealizing is a large part of the explanation of why they are not extensionally equivalent to Act Consequentialism. The idealizing evaluation of rules helped to explain why Rule Consequentialism does not collapse into Act Consequentialism. Given that the relevant kind of idealization is incompatible with the willingness requirement, we might wonder about the practical significance of a theory of pattern-based reasons that accepts any form of this requirement.

5.6 Narrowing Eligibility

In fact, depending on what else we say about eligibility, it could turn out that there is a pattern-based reason to do *X* only if there is an act-based reason to do *X*. To see how this could happen, consider a case in which the agent faces a choice between keeping a promise or failing to keep it. Let us say that there is a pattern of action, *P*, which consists of others keeping their promises, plus this agent keeping his promise on this occasion. Could it be that there is a pattern-based reason for him to keep this promise, but not an act-based reason to do so? If so, keeping it on this occasion would have to have worse consequences than breaking it would. But if that is the case, why should we think that he has a pattern-based reason to keep his promise? After all, there would seem to be an alternative pattern, which we can call *P**, which consists of everyone else keeping their promise plus our agent breaking his, on this occasion. We have just noted that this pattern must have better consequences than *P*—otherwise he would have an act-based reason to keep his promise. So, his strongest pattern-based reason is to play his part in *P**, and his part in this pattern is to break his promise.⁵¹ The general problem is to see how it is possible to have a pattern-based reason to play one's part in a pattern without it also being true that one has an act-based reason to do the very same thing.

⁵¹ See Regan 1980: 54–5.

If there is an act-based reason to do X whenever there is a pattern-based reason to do X , as this argument suggests, it would not follow that pattern-based reasons have no ethical significance. First, they would retain one kind of significance by resolving some indeterminacy in pure coordination problems, as we noted when discussing the ‘collapse’ objection to Rule Consequentialism. Second, even though pattern-based reasons would favour only those actions that are also favoured by act-based reasons, they would provide a different kind of reason for these actions. Each agent’s reason would be ‘this is my part’, not ‘this makes the outcome best’. As I have noted, this difference may be important to us. So far as the agent’s act-based reasons go, his actions are just one possible cause of good outcomes, like all other causes. So far as his pattern-based reasons go, his actions are uniquely significant: he cannot perform anyone else’s part in valuable patterns. For some purposes, this difference may matter even if the favoured actions are exactly the same.

However, this still does not get us what we wanted, which is an explanation of moral constraints. For that, we would have to find some problem in the argument we just considered for the conclusion that there is an act-based reason to do X whenever there is a pattern-based reason to do X . I have already noted that idealizing accounts of the eligibility of patterns are one way to avoid this conclusion: they provide grounds for denying that P^* must have better consequences than P , according to some preferred (idealizing) account of ‘the consequences of P ’. But if we reject idealizing views, is there any way to avoid the problematic conclusion?

There is at least one way. An essential premise in the argument was that there must be an eligible pattern P^* which consists of others performing their parts in P plus the agent failing to perform his part in P . In our example, P^* is the pattern consisting of others keeping their promises plus the agent breaking his promise on this occasion. Now, this is a possible pattern of action, in our technical sense of ‘pattern’ according to which any arbitrary combination of token actions is a pattern. But in an ordinary sense of ‘pattern’, I am not conforming to the ‘pattern of promise-keeping’ if I break my promise.⁵² In this ordinary sense, not all arbitrary combinations are patterns: a pattern is more akin to a ‘practice’ or ‘established activity’.

Perhaps, then, the constraints on the eligibility of patterns are *more restrictive* than we have been assuming. Suppose, for instance, that only

⁵² Rawls 1955: 17 claims that the practice of promise-keeping specifically excludes making an exception in order to promote the good. This seems to me an exaggeration. In any case, the present point is merely that, in an ordinary sense of ‘pattern’, one is not complying with a pattern of promise-keeping if one breaks a promise.

those patterns which constitute practices are eligible. If that were true, we should not expect there to be an act-based reason to do *X* whenever there is a pattern-based reason to do *X*. Sometimes, as in the promise-keeping example, more could be achieved by departing from the practice, so there would be an act-based reason to do that, but there would be a pattern-based reason to conform to the practice—since there would be no superior eligible pattern than the one constituting the practice.

In fact, the idea of a practice may help us to explain Jim's reason not to shoot, even while accepting the willingness requirement. The key to this is to shift focus from Pedro to other potential cooperators. Though Pedro is unwilling to play his part in the pattern <neither Jim nor Pedro shoots anyone>, perhaps this is not the pattern that explains Jim's reason not to shoot. Another relevant pattern, which many people cooperate in realizing, is <rights respecters not violating others' moral rights>.⁵³ If this pattern of respect for moral rights constitutes a practice, it might be that Jim has a pattern-based reason associated with it. Moreover, there is no practice consisting of others respecting moral rights and Jim shooting an innocent person; so there is no danger that this pattern will turn out to be eligible and superior, on the view we are entertaining.

The practice view looks promising, but it is not entirely satisfactory. One problem with it is that it seems too narrow. In particular, we almost certainly want to treat some patterns that do not constitute practices as eligible. This is true, for example, of many of the coordination cases discussed by Gibbard and Regan. It may also be true in cases where we want to account for the emergence of novel practices, and we want to point to pattern-based reasons for getting a practice started.⁵⁴ Another problem is that we have not offered any rationale for the restriction to practices, other than that it is a way of avoiding the conclusion that there is an act-based reason to do *X* whenever there is a pattern-based reason to do *X*.

⁵³ I shall discuss how to understand this pattern in Chapter 6. It might be that there is more than one relevant pattern, of course. For example, in not shooting Jim might also participate in the pattern <respecters of the right not to be killed arbitrarily respecting the right not to be killed arbitrarily>. There might even be many such patterns. A full account of which patterns are eligible in cases like this would require a full theory of eligibility, which I do not aim to provide here. However, the form of pluralism outlined here is open to the possibility of there being more than one pattern-based reason for performing a single action, where that action is the agent's part in more than one eligible pattern. I am grateful to Tim Mulgan for comments on this point and for the suggested alternative pattern.

⁵⁴ Scanlon 1998: ch. 7 criticizes practice accounts of reasons to keep promises. Kolodny and Wallace 2003 defend a practice account from Scanlon's criticisms. Even if they are right about promises, though, we should not expect all pattern-based reasons to be associated with practices.

So we should be clear about what has and has not been shown. I have agreed with Dietz that, if we believe that there are pattern-based reasons, we should accept a weak form of willingness requirement. This is to accept that a pattern is not eligible unless there is a sufficient chance that it would be realized, were the agent to play his part in it. This means that our theory of pattern-based reasons will not be idealizing, unlike standard forms of Rule Consequentialism. This raised once again the issue of whether pattern-based reasons have much practical significance. Though they can supplement act-based reasons in coordination problems, they will not conflict with them if the only constraints on eligibility are the minimal constraint and the weak willingness requirement.

However, conflict between act-based and pattern-based reasons is possible if we accept further constraints on eligibility. We can see this by supposing that only those patterns which constitute practices are eligible. If that were true, there could certainly be conflicts between act-based and pattern-based reasons, of the sort that may help to explain our convictions about cases like Jim's, in which a moral constraint seems to be in play.

These conclusions set an agenda for further development of the theory of pattern-based reasons. If we accept Dietz's argument, and the explanatory hypothesis about the strength of pattern-based reasons on which it rests, we should look in a different direction than Rule Consequentialism. Rather than accepting only the minimal constraint on eligibility, we should look for *further* constraints on eligibility, in addition to a weak willingness requirement. These constraints should imply that the patterns which constitute practices are eligible, together with some other patterns.⁵⁵ Furthermore, we should look for some rationale for why these patterns, and not others, give rise to pattern-based reasons.⁵⁶

⁵⁵ One possibility is a stronger form of willingness requirement, according to which a pattern is eligible only if others are 'cooperative' in the stronger sense that they would behave in the requisite way *with certain kinds of intention*. It might be possible to account in this way both for practices and for novel acts of cooperation.

⁵⁶ A further important question is whether there is a pattern-based reason to play one's part in *every* good eligible pattern, or only the *best* eligible patterns. The second option may initially seem attractive. To develop it, we would need to specify the relevant comparison class—but perhaps that would not be too hard: for any pattern *P* the comparison class might be the set of all eligible patterns performable by the same set of agents on the same occasion. However, I am inclined to think that the first option is correct: every good eligible pattern generates a pattern-based reason to participate; however, we may be justified in ignoring those which are not best, since they are not likely to give rise to the strongest reasons. I am inclined also to say the same thing about act-based reasons: I have some reason to perform any good option, even though I will have a stronger reason to perform any better alternative to it.

5.7 Conclusion

Those inclined to think that Act Utilitarianism is too simple should consider the merits of adopting a more complex structure. This is an alternative to sticking with Act Consequentialism but adopting a more complex theory of value.

Adding complexity to the structure of Act Utilitarianism means departing from Act Consequentialism. I have argued that any plausible theory of reasons will claim that we have act-based reasons on all the occasions when Act Consequentialism claims that we have them. In that sense, Act Consequentialism is at least half right. However, we may also have other reasons, which derive from the fact that actions may be parts of ethically significant patterns. The concept of pattern-based reasons, I claim, promises a better explanation of moral constraints than is possible in any form of Act Consequentialism. Constraints single-out the agent's action, so far as the agent's reasons are concerned. If we stick with Act Consequentialism we can explain this only by making very implausible claims about value. In contrast, the notion of playing one's part in a pattern already singles-out the agent's action.

However, departing from Act Consequentialism need not mean embracing Rule Consequentialism. Dietz's arguments show that it is at least very hard to give a satisfactory account of the strength of pattern-based reasons without accepting a version of the willingness requirement. This strongly suggests that a satisfactory theory of pattern-based reasons will not be idealizing in the way that Rule Consequentialism is idealizing. If that is correct, and pattern-based reasons are to be capable of explaining moral constraints, then there must be additional constraints on eligibility over and above the minimal constraint and the weak willingness requirement. These constraints should allow that the patterns which constitute practices are eligible, along with some others.

I will not pursue these difficult questions about eligibility any further in this book. Instead, I will try to show that the concept of pattern-based reasons can be employed in promising ways, including in explaining moral rights. To make these claims I will assume that certain specified patterns are eligible. If the explanations these assumptions afford do seem promising, that may provide motivation to renew the search for a fully satisfying account of the eligibility of patterns.

6

Moral Rights

Much moral and political discourse employs the concept of moral rights. Persons, and sometimes other entities, are claimed to have moral rights that protect them against harm or interference, or entitle them to certain kinds of assistance or goods. These moral rights are supposed to be distinct from legal rights, and to be in conflict with them in certain cases. Indeed, one context in which moral rights seem especially salient is when existing laws do not assign legal rights corresponding to the moral rights which, we think, someone has. When the law restricts or denies legal entitlements to education, for example, we may protest by claiming the existence of a ‘right to education’. The most natural interpretation of this claim is that it is asserting the existence of a moral right to education.

One important question for utilitarians is whether to offer a positive theory of moral rights—that is, a theory which seeks to account for the existence of moral rights. Bentham thought that talk of moral rights was confused and politically dangerous, famously decrying it as ‘nonsense upon stilts’.¹ We do not have to share Bentham’s background assumptions about language and morality to wonder whether we should also try to do without the concept of moral rights. There are significant difficulties in developing a satisfactory theory of moral rights—and not just for utilitarians. Moreover, we can interpret at least some uses of the language of moral rights in other ways. For example, we can interpret some claims to the effect that a moral right exists as, strictly speaking, claims that a legal right ought to exist. Following this suggestion, we might interpret ‘there is a right to education’ as claiming that the law ought to be changed so as to include such a right.

However, this sort of manoeuvre does not capture the content of much ordinary moral and political thought about moral rights. We tend to think that if someone is detained without a fair trial in a state that recognizes no

¹ I am here treating ‘natural right’ and ‘moral right’ as synonyms. On Bentham’s views about natural rights see Schofield 2006: ch. 3.

legal right to a fair trial, that person has been wronged because his moral right to a fair trial has been violated. We cannot capture this thought by saying merely that the law in this state should be changed, since that does not entail anything about the rightness or wrongness, as things stand, of this person's detention. Nor, it seems, can we capture the thought that he has been wronged merely by saying that his detention was morally wrong. The detention was not only wrong: it was also, in some sense, 'directed' at him. One way to vindicate the common sense distinction between (merely) *acting wrongly* and (in addition) *wronging someone*, relies on the concept of moral rights.²

For this reason, we should see how far we can get in developing a utilitarian theory of moral rights. The topic of moral rights is huge and complex, and we should not expect anything like a complete and satisfactory account of it. More modestly, we will try to identify the main problems standing in the way of a utilitarian theory of moral rights, and to explore some strategies for solving those problems.

6.1 The Concept of Moral Rights

To get a fix on the task, let us review briefly some of the features of the concept of moral rights.

First, like legal and customary rights, moral rights (if they exist) appear to be analysable as 'Hohfeldian incidents'. A Hohfeldian incident is one of four atomic kinds of rights, out of which complex 'molecular' rights may be composed. The four atomic kinds are: privilege, claim, power, and immunity.³ A *privilege* corresponds with the absence of a duty on the part of the right-holder: you have a right to scratch your own nose in this sense, since (we may assume) you have no duty not to scratch it. A *claim*, meanwhile, corresponds with some other agent's duty: your claim that I do what I promised corresponds with my duty to do what I promised. A *power* is an ability to change someone's privileges or claims. For example, you have the right to promise, which is an ability to create a claim on the part of the person to whom you promise that you do what you promise; the promisee will in turn acquire a power to release you from your promise. Finally, an *immunity* corresponds with some other agent's lack of (Hohfeldian) power.

² On 'directed duties' see Cruft 2013. Nelson 2015 claims that utilitarianism cannot account for them.

³ See Hohfeld 1913 and the excellent discussion in Wenar 2005.

For example, constitutionally enshrined rights grant immunities, insofar as they deny legislatures the power to alter the basic rights of citizens.

This analysis appears to apply to rights in general. If you are queuing for a bus, you have a customary claim right that anyone not in front of you in the queue not try to board the bus before you do. If the law grants you a property right in your pencil, this consists in a bundle of privileges (for example, you have no legal duty not to use it to write, and no legal duty not to use it as kindling), plus a bundle of claims (for example, one against each other person that he or she not use it without your permission), plus a bundle of powers (for example, to give it to someone else), plus a bundle of immunities (for example, others lack the legal power to sell it without your permission). There seems no reason to doubt that moral rights, too, should they exist, would be analysable in this way.

Leif Wenar has emphasized the variety of functions performed by rights. We may be tempted to adopt some simple theory, such as the theory that all rights protect the rights-holder's interests, or the rival theory that all rights protect the rights-holder's discretion. Each of these simple theories of the functions of rights seems to apply to some rights, but not all. A very young child's (claim) right not to be abused protects his interests but not his discretion. On the other hand, an army officer's (power) right to order his soldiers to attack protects his discretion but not his interests. Meanwhile, in a mandatory sentencing regime the (power) right of a judge to sentence a criminal protects neither the judge's interests nor her discretion. Wenar argues persuasively that we should accept that different rights have different functions.⁴

Despite this variety in kinds of rights and the functions of rights, Hohfeld's analysis suggests that all rights correlate with some combination of 'duties' (or lack of duty) and 'abilities' (or inabilities) to change duties. How should we understand 'duty' in this context? One option is to understand it as a synonym for 'obligation' or 'requirement'. This would entail that, for example, it could never be legally right to infringe a legal right, nor morally right to infringe a moral right. On the assumption that acting rightly is always possible, this in turn would mean that rights cannot conflict with each other—for if they could, situations could arise in which it would be

⁴ Wenar 2005: 237–52. The examples are drawn from Wenar's discussion. Defences of the 'interest theory' include Lyons 1994: ch. 1 and Raz 1984. Defences of the 'will theory' include Hart 1982: ch. 7 and Steiner 1994.

impossible to infringe no rights at all, and so impossible to act rightly. This would be a substantial constraint on what rights could exist.⁵

The alternative is to treat ‘duty’ as a synonym for ‘pro tanto reason for action’. This would entail that, if you have a (claim) right that I lend you my hammer, I have a reason to lend you my hammer; but it could nevertheless be that, all things considered, it would be wrong for me to lend it to you. We could then distinguish between right and wrong infringements of rights, reserving ‘violation’ for wrongful infringements.⁶ Since, it seems to me, common sense distinguishes between right and wrong infringements of rights and allows for conflicts of rights, this is the understanding of ‘duty’ that I will employ.

This means that we propose to understand rights as ‘correlating’ with reasons and with abilities (or inabilities) to change reasons. The right not to be tortured correlates with a reason not to torture, and the right to make a promise correlates with the ability to create a reason to keep the promise made. What might explain the correlation between rights, reasons, and abilities to change reasons? One possibility is suggested by Joseph Raz: we could think that rights are ‘grounds’ of reasons.⁷ One attractive feature of this suggestion is that, if it were correct, we could explain reasons in terms of rights. For example, we could explain the reason not to torture as grounded in the right not to be tortured. This is, on the face of it, an appealing kind of explanation.⁸ On the other hand, making good on Raz’s suggestion would require that we explain what rights are in a way that distinguishes them from the reasons they ground. In the case of legal rights, we could perhaps do this by pointing to pieces of law—but it is harder to see how to do it in the case of moral rights.⁹

A simpler explanation of the correlation between rights, reasons, and abilities to change reasons is that rights simply *are* structures of reasons and abilities to change reasons.¹⁰ On this view, the right not to be tortured does

⁵ Steiner 1994 argues from this constraint to the claim that all rights are property rights.

⁶ This terminology is due to Thomson 1990: 122. Oberdiek 2004 criticizes the distinction and what he calls the ‘moral space conception of rights’ (327).

⁷ Raz 1984: 196–7. Similarly, Feinberg claims that ‘claim-rights are somehow prior to, or more basic than, the duties with which they are necessarily correlated’ 1970: 250.

⁸ Frances Kamm offers this sort of explanation of constraints. See Kamm 2007: 29.

⁹ Note that I am claiming that Raz’s view carries an extra explanatory burden, not that it could not be discharged.

¹⁰ This may be Wenar’s view, since he claims that ‘[a]ll rights are Hohfeldian incidents’ 2005: 252. The ability to change or create reasons can seem mysterious: like a bit of normative magic. But as David Owens points out, it is not. Since reasons depend on features of the agent’s circumstances, we should expect the ability to change those circumstances to yield an ability to

not ground the reason not to torture: it *is* that reason (or rather, it is a constellation of reasons not to torture, one for each agent). Similarly, on this view the right to make a promise just is the ability to create reasons to keep promises made. The attraction of this view is that it is simpler. If we can account for the reasons and abilities correlated with rights, we get the rights themselves for free—since rights just are structures of reasons and abilities to create or change reasons. Of course, this proposal alters the sorts of explanations we can give. We can no longer say, with Raz, that it is possible to explain reasons in terms of rights. On the other hand, we can still seek to explain the rightness of actions in terms of rights, since that is to explain the rightness of actions in terms of the reasons that rights consist in.

Let us explore this way of thinking of moral rights, and see whether utilitarianism is able to give a plausible account of the reasons and abilities to change reasons that, we are supposing, rights consist in. In order to do that it will have to be able to explain reasons of the correct sort. Something would not count as a right in the usual sense unless the reasons not to infringe it, or the reasons it enables the agent to create, have certain features. Three such features seem especially important.

First, many reasons associated with rights are *agent-relative*, in the sense that they are reasons for each agent not to infringe the right herself, rather than reasons to see to it that infringement does not happen.¹¹ If you have a claim against me—say, because we made a deal, and you have kept your part—this means that I have a reason, which others lack, to keep my part of the deal. This reason is relative to me in the strong sense that it is a reason to do something that only I can do.

Second, according to our ordinary concept of rights, reasons to respect rights sometimes act as *constraints*. That is, they sometimes set moral limits to promotion of good outcomes. If there is a moral right not to be tortured, then it is sometimes wrong to torture when doing so would make the outcome better. Similarly, if someone has a moral right to free speech, it is

create or change reasons—and it does. In Owens' example, '[b]y breaking his arm I can make it the case that I am obliged to take this man to hospital and thereby ensure that I am justified in skipping today's class' Owens 2012: 249. This is not to say that there is nothing puzzling about our normative powers: a fuller utilitarian account of rights would say more about powers than I do here. For discussion see Owens 2012.

¹¹ See Pettit 1988: 46. There are agent-neutral reasons to respect rights, too. For example, everyone has a reason to do what they can to increase general respect for rights. But it is the agent-relative reasons to respect rights that seem trickier for the utilitarian to explain.

sometimes wrong to hinder her speech when doing so would promote the good.¹² These constraints need not be absolute: it could be that it is sometimes morally right to infringe someone's right for the sake of promoting the good. But if there are moral rights they will set some limits to promoting the good, and in those cases where the right ought to be infringed the reason associated with the right will have been outweighed. The reason to respect the right is not itself conditional on respect for the right on that occasion also serving to promote the good.

Third, reasons associated with rights are also, according to our ordinary concept of rights, *usually decisive*, in two ways. They usually outweigh reasons with which they come into conflict in determining which actions are right, and virtuous agents usually treat them as decisive in their deliberations.¹³ This is a watered-down version of Ronald Dworkin's claim that rights are 'trumps'.¹⁴ Dworkin's claim suggests that reasons associated with rights can never be outweighed by reasons of other kinds. That seems an exaggeration of our ordinary understanding of rights; many people think that there are rights, but that they are sometimes defeated—for example, by considerations of public safety or well-being. However, the more moderate claim that they are usually decisive does seem to reflect a feature of our ordinary concept of rights. If someone were to assert that there are rights but that the reasons associated with them are easily outweighed, we would take them to be using the concept in a revisionary way.

Reasons associated with rights are often agent-relative, they often act as constraints, and they are usually decisive. These seem to be essential components of our ordinary concept of moral rights. Can utilitarians account for them?

6.2 Existing Utilitarian Theories of Moral Rights

Some forms of consequentialism incorporate respect for rights directly into their theory of value. That is, they treat an instance of rights fulfilment as a

¹² Dworkin takes this feature of rights to be incompatible with utilitarianism: 'when we say that someone has a right to speak his mind freely... we mean that he is entitled to do so even if this would not be in the general interest. If we want to defend individual rights in the sense in which we claim them, then we must try to discover something beyond utility that argues for these rights' 1978: 271. Dworkin's claim assumes that utilitarians must be Act Utilitarians.

¹³ See Brink 2013: 215–16. My discussion of rights is influenced by Brink's work.

¹⁴ Dworkin 1984: 153. See also Pettit 1988: 46.

noninstrumentally good feature of outcomes, and an instance of rights infringement as a noninstrumentally bad feature of outcomes.¹⁵ It may seem as though this claim cannot be made by utilitarians, because of their commitment to welfarism. Strictly speaking, utilitarians can make this claim, so long as they treat instances of rights fulfilment as themselves constituents of well-being. They can then say that respecting someone's right has noninstrumental value, since it contributes directly (as a constituent) to her well-being.

Whatever the merits of 'goal rights' consequentialism, though, this strategy does not look very promising for utilitarians. Apart from anything else, it is not very plausible to think of instances of rights fulfilment as constituents of well-being. It is much more plausible to think of instances of rights fulfilment as, at least some of the time, causes of well-being. A right to free speech, for example, may be said to protect important interests of agents in expressing and hearing ideas. It is not very plausible to claim in addition that fulfilment of this right is a constituent of well-being. Rights seem to contribute to well-being, so far as they do, via what they enable or protect. So the more promising strategy is to claim that rights have some sort of instrumental value. An immediate objection to that way of thinking of them, however, is that respecting rights quite often seems not to be beneficial. Indeed, we earlier claimed that an essential feature of the ordinary concept of rights is that the reasons associated with them often act as constraints on promotion of the good.

There are a number of ways for utilitarians to try to account for this feature of rights. One is to distinguish kinds of well-being, and to assign much greater value to the kinds that rights are said to protect. As David Brink explains, this is one way to interpret John Stuart Mill's claims about rights, and it is consistent with the interpretation of Mill as an Act Utilitarian.¹⁶ On this interpretation, Mill claims that rights protect deliberative capacities, and that deliberative capacities are the 'pre-eminent' constituent of well-being. If deliberative capacities are lexically prior in importance to other constituents of well-being, then even an Act Utilitarian should treat respect for rights as a constraint—not on pursuit of the good as such, but on the pursuit of other constituents of well-being.¹⁷

¹⁵ See Sen 1982. For criticism of this sort of view, see Nozick 1974: 28–9 and Foot 1985: 197–8.

¹⁶ See Brink 2013: 226–32.

¹⁷ Good *A* is lexically prior to good *B* if and only if no gain in *B*, no matter how large, would compensate for any loss in *A*, no matter how small, and any gain in *A* would compensate for any loss in *B*. See Rawls 1999a: 37–9—though note that he characterizes lexical priority as a relation among principles, not among goods. He also notes that the concept of lexical priority 'appears to offend our sense of moderation and good judgment' (38).

Arguably this does not fully capture the sense in which, according to our ordinary concept, reasons to respect rights are constraints. In any case, this way of trying to reconcile Act Utilitarianism with a theory of rights cannot capture one of the other features we are assuming is essential—namely that reasons associated with rights are often agent-relative. Act Utilitarianism claims that each of us is required to maximize well-being, and assigns to each of us only agent-neutral reasons. If some constituents of well-being are pre-eminent, Act Utilitarianism implies that we have agent-neutral reasons to promote them. Thus, it implies that each of us should see to it that rights are respected, but it does not imply that we have agent-relative reasons to respect them ourselves. On this Act Utilitarian view, each of us should violate a right every time that would maximize the good—which is inconsistent with our common conception of reasons to respect rights.

A second kind of utilitarian theory of rights says that their moral significance is captured by the fact that respect for them is part of the correct decision procedure. On this view, the best psychology for utilitarian agents will include a disposition to follow a rule of the form ‘respect rights *X*, *Y*, and *Z*’, where ‘*X*, *Y*, and *Z*’ is some suitable list of rights. This sort of utilitarian claims that this psychological disposition promotes utility better than other feasible dispositions. According to this view, the moral significance of rights is that respect for them is part of the psychology that it is best for agents to have.¹⁸

It is important to note that the disposition could be strict or even exceptionless: it need not involve treating respect for rights as a mere ‘rule of thumb’. The reason for this is that being disposed to depart from the rule when doing so would seem to have better consequences may itself (as a disposition) have worse consequences than being disposed more strictly to follow the rule. Spotting exceptions to the rule is costly and highly fallible. Being on the lookout for exceptions may also be incompatible with important goods that depend on secure expectations of others’ behaviour.¹⁹ If the best psychology includes a strict or only rarely defeasible disposition to respect others’ rights, this sort of view will be able to account for one of the senses in which, according to our ordinary concept of rights, they are

¹⁸ Theories of this sort include Hare 1989: chs. 7–9 and Pettit 1988. Brink 2013: 223–6 notes that Mill can be interpreted as offering this sort of theory. Gibbard 1984 and Arneson 2001 also offer theories along these lines, though they focus on which dispositions should be taught or advocated.

¹⁹ Pettit 1988 argues that believing others to be on the lookout for exceptions to rules requiring respect for rights is incompatible with having the important good of dignity. As he notes, some other goods—including pleasure and spontaneity—also cannot be ‘deliberatively internalised without self-defeat’ (50).

usually decisive: virtuous agents will usually treat them as such in their deliberations.²⁰

These claims about the best decision procedure for utilitarian agents are certainly important. However, by themselves they do not account for the full moral significance that we usually take rights to have. They do not ascribe any significance to rights so far as our (non-perspectival) reasons go, nor in determining the (non-perspectival) rightness of actions.²¹ They suggest that, if you could maximize utility by torturing me, the best way for you to decide what to do would be to act on a strict or even exceptionless disposition not to torture people. But they do nothing to explain the conviction that I have a right not to be tortured that is correlated with a reason that constrains your promotion of the good, nor the conviction that torturing me would be wrong in part because it violates my right.

We cannot explain all three essential features of the ordinary concept of rights by making claims about value, nor by making claims about the best decision procedure, nor by simply combining both sorts of claim. If we want a utilitarian theory of rights which implies that agents have usually decisive and agent-relative reasons that act as constraints on promotion of good outcomes, we need to depart from Act Utilitarianism and pursue some version of the third strategy utilitarians can adopt to explain rights. That is, we need to adopt some indirect utilitarian account of reasons. One example of an indirect account is the sanction theory of rights apparently endorsed by Mill. According to this theory, you have a right to *X* if and only if the social practice of sanctioning (punishing) any violations of *X* would maximize utility.²² This theory is indirect in the sense that it seeks to account for the right—and presumably, the reasons correlated with it—in terms of the utility of something other than exercise of or respect for the right. It is the utility of the social practice of sanctioning violations that is said to explain the right.²³

²⁰ On the other hand, if the best decision procedure for an agent depends on her cognitive abilities, some agents might be much less constrained than others by rights. This conflicts with the idea that moral rights place all agents under the same sorts of constraint.

²¹ For this reason, Frey claims that moral rights are at most 'mere appendages' for Act Utilitarians 1984: 66.

²² See Brink 2013: 217–23. See Mill 1998/1861: ch. 5, where Mill writes: '[w]hen we call anything a person's right, we mean that he has a valid claim on society to protect him in the possession of it . . . To have a right, then, is, I conceive, to have something which society ought to defend me in the possession of' (97–8).

²³ My statement of the sanction theory follows Brink's. Strictly, the statement claims only that rights and utility-maximizing sanctions are coextensive; it does not make any explanatory claim. However, the correlation would be mysterious without some explanatory claim, and this is the only one that seems available to the utilitarian.

The sanction theory is, in one way, attractive. By shifting the focus of evaluation from individual acts of respect for rights to a social practice, it offers the prospect of retaining a broadly instrumental conception of rights while being able to explain how there can be reasons to respect rights even when doing so does not maximize utility.²⁴ Moreover, it is quite attractive to think that the social practice of respect for rights is instrumentally valuable. We can see this by comparing the practice of respect for rights with the practice of punishing people. Even those who reject instrumental approaches to justifying individual acts of punishment often accept that the justification of the social practice of punishment is instrumental or utilitarian in character.²⁵ If we came to think that the practice of punishing people failed to promote well-being overall, many of us would conclude that the practice should be either reformed so as to promote well-being, or abolished. I take it that the parallel view about rights is similarly plausible. If we came to think that our practice of respecting each other's rights failed to promote well-being overall, many of us would conclude that the practice should be either reformed or abolished.

However, the sanction theory itself is not satisfactory. One reason is that, as Brink points out, it seems to get things backwards. The sanction theory says that you have a right to X because it is appropriate to sanction violations of it, whereas we are more likely to think that the violation of the right explains the appropriateness of sanctioning the violation.²⁶ More importantly, the sanction theory seems too narrow. It is, as we have noted, a version of the idea that rights are to be explained in terms of the utility of social practices of respect for rights. But these practices include many more things than sanctioning violations. As well as sanctioning violations, we take rights

²⁴ David Lyons has been an influential critic of indirect utilitarian theories of rights. He asks, '[i]f a utilitarian believes that [legal rights-ascribing] rules are justified on utilitarian grounds, does he *contradict* himself by supposing that direct utilitarian arguments for deviating from the rules may be entertained? I see no contradiction here; and in that case, the utilitarian cannot understand the legal impact of such rules automatically to translate into moral force, even when those rules are supported by the best arguments. He cannot regard the morally defensible rights under *utilitarian* institutions as having moral force' 1994: 168, emphases in the original. Lyons seems to reason as follows: a utilitarian cannot coherently deny that there can be act-based reasons to deviate from the best rules; therefore, the best rules have no 'moral force'. But this assumes that, unless the best rules are such that there is never a reason to deviate from them, those rules have no moral force. No reason is given for this assumption, which in any case seems clearly false. For further critical discussion of Lyons, see Hare 1989: ch. 8 and Brandt 1992: ch. 11.

²⁵ See Rawls 1955 and Hart 1959–60.

²⁶ Brink 2013: 223. A similar point is made in Hart 1982: 92–3. I treat this reason as less important than the following one, because in general we should allow that philosophical explanations can run in unexpected directions.

into account when deliberating, we use them to form expectations about behaviour, we apologize for failures to respect them, and so on. If our general strategy is to point to the utility of practices of respect for rights, it is not clear why we should ignore everything but the sanctioning elements of those practices.²⁷

6.3 A Broader Indirect Theory

The sanction theory seems too narrow. In the remainder of this chapter we will explore the prospects of a broader indirect theory of rights.

The basic idea of the broader theory is that we think of moral rights as together comprising a beneficial scheme, and of the scheme as being 'embedded' by being respected.²⁸ Respect for moral rights is then a good pattern of action. Supposing that it is also eligible, each of us has pattern-based reasons for playing our part in it, which consists of respecting moral rights ourselves. These reasons are therefore agent-relative, since they are reasons to do something each person alone can do: to respect rights herself or himself. They are not reasons to maintain the scheme, but instead to participate in the beneficial pattern of respect for the rights the scheme consists in.²⁹

To develop this theory we need to say more about the component ideas. The theory depends on the idea that there is some scheme of moral rights, *S*, which satisfies the following two conditions. First, it must consist of rights that are sufficiently close to the moral rights that, intuitively, we believe exist. If it does not meet this condition, the theory will not be able to explain moral rights as we understand them—it will at best be a theory of a different phenomenon. Second, it must be a scheme that a sufficient number of people are already respecting. Except in very unusual conditions, this is necessary to meet the weak version of the willingness requirement that we adopted in Chapter 5.³⁰ We can call this pattern of respect '*P*'.

²⁷ The breadth of the sanction theory of rights depends on our conception of sanctions: it may be that Mill had a broad conception of sanctions, which includes some of the things I claim are left out by the sanction theory. I do not mean to take a stand on this issue of Mill scholarship.

²⁸ On embedding see Chapter 3, Section 3.5, this volume.

²⁹ The present aim is not to explain all of the reasons associated with rights—which might include act-based or pattern-based reasons to establish a different practice of respect for rights, for example. Instead the aim is to show what would be involved in explaining the reasons constitutive of moral rights as pattern-based reasons, and to suggest that this is possible.

³⁰ This requirement is the key difference between the theory proposed here and Rule Utilitarian theories of moral rights such as those discussed in Brandt 1992: chs. 10–11 or Harsanyi 1980.

These two conditions are in tension with each other, in the sense that satisfying one makes it harder to satisfy the other. In fact, we cannot assume that there is any scheme that satisfies both conditions at once. In particular, it is not plausible to treat the content of *S* as settled by the content of local laws or customs. In most cases that would be a way of meeting the second condition, since in most cases there is some respect for local laws and customs. But it would not meet the first condition, since one platitude about moral rights is that they are sometimes out of step with local laws and customs. We want to be able to say, for example, that there is a moral right not to be tortured even in jurisdictions where torture is legally permitted.

Though there is no guarantee that any scheme satisfies both conditions, we can expect that a scheme with the following two additional properties would be more likely to do so than a scheme without one or both. A scheme is more likely to satisfy our two conditions if its content is more abstract and limited. Take, for example, a right to free expression. We can elaborate this right in more specific ways, which differ from each other. One specification of the right may include a right not to be subject to surveillance by security services, while another excludes this. Obviously, in general the number of people respecting either of the more specific rights will be smaller than the number respecting the more abstract or indeterminate right. As we make rights more specific we tend to narrow the base of existing respect for them. So we should expect that, if any scheme satisfies the conditions on *S*, it is likely to contain relatively abstract and indeterminate rights.³¹

The second property that makes it more likely that the two conditions on *S* can be jointly satisfied is that the content of *S* is such that *P*, the pattern of respect for the rights *S* contains, is a beneficial pattern of action across some relatively wide range of levels of compliance with that pattern. If this is the case, pattern-based reasons associated with this pattern will not typically come into and go out of existence as the number of people participating in the pattern changes. So long as some others are respecting the rights contained in *S*, there will be a reason to play one's part in the beneficial pattern of respect they have established.

If there is a scheme of rights with these properties, it stands a better chance of satisfying the two conditions we identified. It will contain rights

³¹ This offers one possible explanation of why 'human rights'—which, on one interpretation, are a category of moral rights—are typically relatively indeterminate. Of course, there are downsides to indeterminacy, as James Griffin emphasizes: see Griffin 2008: 14–18.

that are relatively abstract and indeterminate, so that it is more likely to be true that sufficient people respect them in all societies.³² And it will contain rights which are such that a pattern of respect for them is beneficial whether relatively few or many others are respecting them. In the next section we will consider whether it is plausible to think that moral rights satisfy these conditions.

Before we get to that, though, we must address a possible objection. We are considering the idea that rights are constituted by reasons that are, in large part, pattern-based reasons to participate in some valuable eligible pattern, *P*. This pattern, *P*, is a pattern of 'respect for the rights in some scheme', *S*. But these rights are the very rights whose existence we are trying to explain. How can their existence be explained in terms of reasons to respect them?

The solution is to be more precise. Suppose, for illustrative purposes, that we believe that there are just two moral rights: a right to freedom of speech, and a right not to be tortured. We can then specify the behavioural and motivational components of *P* without referring to any rights or reasons. For example, we can say that *P* consists of people not interfering with each other's speech, not torturing each other, treating the fact that some action *X* would be an instance of interference with speech or of torture as usually sufficient to conclude deliberation whether to *X*, and other similar facts.³³ To participate in *P* is to adopt similar behaviour and motivations. Now, we have specified *P* independently of the reasons and rights we are trying to explain. For example, to treat the fact that *X* would be an instance of interference with speech as usually sufficient to conclude deliberation whether to do *X* is to have a certain psychological disposition. To fail to do *X* is to exhibit certain behaviour. *P*'s existence, strictly speaking, consists only in facts like these, none of which consists in the existence of rights or reasons. To describe *P* as 'respect for the right of freedom of speech and the right not to be tortured' is helpful shorthand, though it is strictly false.

Thus the idea, more precisely expressed, is that there exist patterns of behaviour and motivation consisting of people not interfering with freedom

³² That is, all currently existing societies. One controversial implication of this theory of moral rights is that they have not always existed: see Section 6.5, this chapter.

³³ Pettit writes: 'To recognise a right to *X* must be to do more than just happen, more or less accidentally, to respect it. It must be to respect the right, because it is a right. The consequentialist who claims to recognise a right must not just behave appropriately. He must behave appropriately because he reasons appropriately' Pettit 1988: 48. Brandt 1992: 201 takes a similar view.

of speech, not torturing each other, and so on, and of treating the fact that some action would fall under one of these kinds as usually sufficient to conclude deliberation whether to do it. These patterns of behaviour and motivation are both valuable and eligible, and so they give rise to reasons to participate in them. These reasons are then reasons for each of us not to interfere with freedom of speech, not to torture each other, and so on, while treating the fact that some action would fall under one of these kinds as usually sufficient to conclude deliberation whether to do it. To have a moral right to freedom of speech, or not to be tortured, is for others to have sufficiently weighty reasons of these kinds. There is no bootstrapping—only the appearance of it, due to the initial imprecise expression of the idea.³⁴

On this view, then, ‘respect for rights’ is a matter both of conforming with the behaviour prescribed by the right and of treating rights as usually decisive considerations. To respect a right to freedom of speech is (at least) to refrain from interfering with speech out of concern to treat not infringing it as a consideration that is usually decisive. More broadly, to ‘respect a right’ involves being disposed to treat infringements, by oneself or others, as very serious matters that provide a prima facie case for moral criticism and, where appropriate, attempts to make amends. Our hypothesis is that the benefits of this pattern of behaviour and motivation generate pattern-based reasons to play one’s part in it, and that together with associated abilities these reasons—the reasons to participate in the beneficial pattern—constitute moral rights.

6.4 The Benefits of Respect for Moral Rights

The indirect theory sketched in the previous section depends on the hypothesis that a pattern of respect for moral rights promotes well-being sufficiently to generate strong reasons to play one’s part in it. Do we have any support for this hypothesis?

What effects a pattern of respect for rights has is an empirical matter, and in principle open to scientific study. We have some social scientific evidence that is relevant to it: for example, cross-national studies of the correlation between subjective reports of well-being and levels of respect for human rights, and Amartya Sen’s work on the relationship between legal

³⁴ For a related discussion see Kolodny and Wallace 2003: 124–5. I am here leaving aside the abilities component of rights, since it is not relevant to the objection under discussion.

entitlements and famine.³⁵ But this evidence is only indirectly relevant, since it concerns the relationship between variables that are at best only associated with the variables we are directly interested in—namely well-being and respect for moral rights. So while we should not dismiss out of hand the possibility that social scientific research will provide strong evidence for or against our hypothesis, for the time being we will have to rely on other considerations.

Provided we exercise due caution, we can rely to some degree on what R. M. Hare called ‘common knowledge’ of ‘how the world actually goes’.³⁶ That is, we can rely to some extent on pieces of common sense such as the claim that general respect for the right not to be tortured promotes well-being. Though this seems obviously true, we have to be careful. It is not obvious whether some particular infringements of this right might do more good overall than harm, or whether a readiness to infringe it on the part of some agents in special circumstances might increase overall well-being. We should be alert to the flaws in claims of common sense, and we should be wary of the temptation to resort to ‘just so’ stories when attempting to support the claim that respect for the moral rights we believe exist promotes well-being.³⁷ But we do not need to give up altogether on the attempt to evaluate that claim; we can instead acknowledge the deficiencies of our evidence by being tentative in our conclusions.

It is quite plausible to think both that there are some mechanisms by which respect for moral rights is likely, in general, to promote well-being, and that there will be some cases in which infringing moral rights would have better consequences than respecting them. Given our earlier assumptions about moral rights, this suggests two tasks. First, we should try to catalogue and analyse the mechanisms by which respect for rights usually promotes well-being, seeking to explain their relationship to some different kinds of rights. Second, we should try to explain why, on the assumption that it is sometimes possible to promote well-being by infringing rights, the

³⁵ Diener et al. 1995 presents evidence of a positive correlation between subjective well-being and respect for human rights. Sen 1981 argues that famine is typically caused by collapse in people’s entitlements rather than collapse in food supply. He writes (8): ‘The mesmerizing simplicity of focusing on the ratio of food to population has persistently played an obscuring role over centuries, and continues to plague policy discussions today much as it has deranged anti-famine policies in the past’.

³⁶ Hare 1989: 101.

³⁷ Richard Posner wrote: ‘Utilitarianism . . . seems to base rights of great importance on no firmer ground than an empirical hunch that they promote “happiness.” That hunch cannot be verified by any tools we have or are likely to acquire’ Posner 1979: 116. This is salutary, even though it is an exaggeration.

pattern that generates weighty pattern-based reasons is one in which agents uniformly respect moral rights, rather than one in which exceptions are made where doing so would promote utility. The second task is important because, although we want to allow that all things considered it may be right to infringe rights on some occasions, we do not want those exceptions to be built into the pattern that is said to generate the reasons in which rights consist.

Let us begin by considering some of the ways in which respect for moral rights is likely, in general, to promote well-being.³⁸ As utilitarians have often emphasized, one of the most important effects of respect for rights is security, which itself seems to be very important for well-being.³⁹ There are of course different kinds of security. One kind has to do with regularity of everyday activities—for example, having meals at set times or other kinds of personal habits. Some find this comforting; others find it boring. For our purposes, two more fundamental kinds of security are much more important.⁴⁰

The first is what we might call personal security. This is a matter of security from attack on one's person. It is plausible to think that this sort of security is among the main beneficial effects of widespread respect for a moral right of bodily integrity and a moral right not to be tortured. Central cases of lack of personal security include physical attack or credible threat of such attack—but we might add that personal security can be undermined by attacks on one's mind as well as by attacks on one's body, as in cases of manipulation, harassment, or some kinds of torture. Unlike some other moral rights, each individual's respect for rights protecting personal security does not need (though it may be aided by) any legal apparatus or coordination with others. In this respect it is a unilateral affair.

³⁸ A fuller treatment of moral rights would consider the question of whether non-human animals have such rights. Here I set that aside. For discussion, see Galvão 2016.

³⁹ See Bentham 1838–43c: ch. 7. Mill wrote: '[f]ew hurts which human beings can sustain are greater, and wound more, than when that on which they habitually and with full assurance relied, fails them in the hour of need' 1998/1861: 104. See also Sidgwick 1907: 439–44 and Bailey 1997: 124–5.

⁴⁰ There are other important kinds of security. One is material security, in the sense of not having to worry about fulfilment of basic needs. Whether there is a moral right to material security is both very important and difficult to decide. One theoretical problem with asserting such a right is explaining on whom the corresponding duties fall: see O'Neill 2005. Another kind of security is what Philip Pettit calls 'non-domination'. In Pettit's sense, to be dominated is to be 'defenselessly susceptible to interference by another' 1996: 576. One kind of protection against domination is widespread respect for moral rights that include Hohfeldian immunities.

The second very important kind of security is security of expectations. Much of our activity requires coordination across time and persons, and for this to be possible it is very important that we have secure expectations of each other's behaviour. We could not construct bridges, roads, or opera houses, nor run labs or universities, nor arrange surprise parties or trade without it.⁴¹ But it is not just mutual enterprise that depends on secure expectations; successfully avoiding each other also depends on it. Some of these functions of security can be achieved through practices of promising and agreement, and respect for the moral right to make promises and for the moral right that agreements are honoured presumably has great benefit. Where the activities are on a large scale and involve strangers, it may be necessary to rely on law, with its coercive sanctions, and not just on respect for moral rights. We shall come back to this shortly, since it offers one way of understanding the complex relationship between moral and legal rights.

A second mechanism by which, it seems likely, respect for moral rights usually promotes well-being is via protection of agents' freedom. One way in which freedom usually promotes well-being is that it allows agents to pursue their own interests and goals. It might be possible to imagine a third party restricting Sally's freedom in such a way as to leave Sally only with options well-suited to pursuit of her goals. But, except in special circumstances, we do not expect this to be an effective way of enabling agents to pursue their goals—third parties usually do not know enough, or are not powerful enough, or not well-intentioned enough. Usually it is better to protect Sally's freedom, and to enable her to make choices in pursuit of her own goals.

But this is by no means the only way in which freedom promotes well-being. It also has considerable epistemic value, both for the agent herself and for others. Being free enables agents to try things out, and perhaps to backtrack—both of which, we claimed in Chapter 4, are likely to be very important in enabling agents to discover what is good for them. This is partly a matter of each agent discovering which things are or are not constituents of her well-being, and partly a matter of her discovering

⁴¹ Hobbes was right to say that without security 'there is no place for industry; because the fruit thereof is uncertain: and consequently no culture of the earth; no navigation, nor use of the commodities that may be imported by sea; no commodious building; no instruments of moving, and removing, such things as require much force' 1946/1651: 82. However, the fact that Hobbes is the arch philosopher of security should alert us to the fact that not every way of achieving security is desirable. One way of securing expectations, for example, is to be ruled by an orderly tyrant. In this respect security is like efficiency: neither is an impressive virtue, though insecurity and inefficiency are serious vices. One way to achieve efficiency, for example, is to allocate all goods to one individual, but this is not desirable. On efficiency, see Rawls 1999a: 59–62.

which ways of combining constituents work well. Moreover, where people enjoy a wide range of options and the opportunity to try things out and to backtrack, and yet they stick with some things rather than others, that also provides some evidence for others that the things they stick with are things they value, and so are things that may be good for them.

A third mechanism by which respect for moral rights is likely, in general, to promote well-being is by distributing authority. Many rights give the right-holder discretion over some matter, and in that sense assign authority to her over whether the thing in question occurs or not. We have just considered one kind of value flowing from this discretion, in protecting freedom. This value goes mainly to the right-holder herself. But distributing authority has an additional kind of value as well, which is more widely shared. In many circumstances it is easier to get things done if individuals are assigned authority and others know where the authority resides. To see this consider the importance, when working for an organization, of coming to know who is in charge of which decisions. Without this knowledge it is very hard to achieve anything; and things would not be better if there were no distribution of authority at all. When authority is distributed and the distribution is known by others it is clear whose agreement is needed for the thing in question to occur. And, on the whole, there are advantages in not merely assigning authority but distributing it widely. Some of those advantages have to do with the value of protecting freedom, but others have to do with guarding against concentrations of power, and others again have to do with the efficiency of division of labour.⁴²

These three ways in which, usually, respect for rights can be expected to promote well-being differ greatly in the extent to which they rely on co-opting other mechanisms. Respecting rights that protect personal security is, we noted, in one way a unilateral affair. Each of us can achieve the relevant good—that is, not subjecting others to personal attack—by acting alone. At the other extreme, securing expectations often involves reliance on legal systems; we secure expectations by treating legal processes as, within limits, authoritative. For example, we make legally binding contracts, and we treat legal property rights as distributions of authority. This fact is of great importance in explaining our moral rights. For it is not true that our moral rights are wholly independent of our legal rights, even though part of their role is as a normative standard for legal rights. One way to gain a

⁴² See Harsanyi 1980: 128.

moral right to something is to make a legally binding contract, and the content of the moral right will in some respects depend on the content of the legal right. Moreover, this can happen even when the legal system concerned is not perfectly just. Suppose that you are legally the owner of your wheelbarrow, and that you agree to lend it to me so long as I fix your gate. When I fix your gate I gain a *moral* right to use of your wheelbarrow—even if the allocation of legal ownership rights in our society is only reasonably, and not perfectly, just.⁴³

The way in which such ‘legitimate expectations’ gives rise to moral rights is, for many theories of moral rights, something of a puzzle. Why should moral rights depend on legal rights? And why should somewhat (though not grossly) unjust legal systems generate moral rights? Using the indirect utilitarian theory of rights, we can explain this as follows. Some of the benefits of respect for moral rights can be achieved only by relying on other mechanisms, and these include legal systems. We can, for example, secure certain kinds of expectations only by treating legal decisions as authoritative. It would not be beneficial, on the whole, to achieve security of expectations by treating grossly unjust legal systems as authoritative, but neither is it likely to be beneficial on the whole to forego the benefits of secure expectations in every case in which a perfectly just legal system is not to hand. Thus we should expect that a pattern of respect for moral rights that is, on the whole, beneficial will include a moral right to treat reasonably just legal systems as authoritative.⁴⁴

No doubt there are other ways in which respect for rights usually promotes utility, and no doubt there are further things to be said about the three mechanisms we have been considering. But let us now move to the second task we identified, of explaining why the relevant pattern of action, for our theory, is one in which agents uniformly respect moral rights, rather than one in which they make exceptions in cases where utility could be maximized by infringing a right. Recall that we are trying to develop a theory of moral rights according to which they consist in pattern-based reasons to participate in some pattern *P*, which itself consists of respect for some scheme of rights *S*. But we are also assuming that, on some occasions, utility would be maximized by infringing one or more of the rights in *S*. In that case, why do we have pattern-based reasons to participate in *P* and not in

⁴³ Rawls’s discussion of ‘legitimate expectations’ is relevant here. See Rawls 1999a: 51–2, 74–6, 273–7, and Chapter 7, this volume.

⁴⁴ This right is, perhaps, one application of a more general moral right to make use of tools.

P^* , where P^* is just like P except that it includes the rights-infringing but utility-maximizing behaviour?

This is a specific version of a general question that we considered in Chapter 5. In this case the answer is that P does not consist only in conforming or infringing behaviour, but also in motivational states.⁴⁵ Moreover, many of the benefits of P flow from the motivational states and not from the behaviour. For example, the strong motivation not to break promises plays an indispensable role in securing expectations. Similarly, the strong motivation not to torture plays an indispensable role in realizing personal security. Well, then, could we keep the motivational states of P fixed, and simply add to them the utility-maximizing behaviour of P^* ? No, because the value of the motivational states depends in part on their being motivationally effective, that is, on causing the behaviour they motivate. We cannot simply mix and match optimific motivation with optimific but incongruent behaviour.

In short, many of the benefits of respect for rights depend on people being motivated by fairly simple *types* of actions and situations. Such motivations help to make behaviour stable and predictable, and—insofar as they are detectable other than by the behaviour they motivate—provide assurance to others. This gives us some reason to think that a pattern of uniform respect for moral rights will have better consequences than one in which agents are disposed to make exceptions whenever doing so seems beneficial.⁴⁶

6.5 The Contingency of Moral Rights

So far we have outlined an indirect utilitarian theory of rights according to which rights consist in abilities and reasons, and these reasons are generated by beneficial patterns of behaviour and motivation. The theory is only in outline, since we have not been able to show that the behaviour and motivation that corresponds with respect for the moral rights we believe

⁴⁵ In this respect the theory developed here resembles Hooker's: see Hooker: 2000: 75–8. However, I am seeking to explain rights in terms of actual rather than ideal patterns of respect.

⁴⁶ Raz 1984: 208–9 suggests that a further function of rights is to provide shared standards in social life—'intermediate conclusions' which command assent even among people who disagree about the underlying reasons. If that is correct, it may be another reason for thinking that simplicity and uniformity in respect for rights is likely to be most beneficial. See also Gibbard 1984: 101–2.

exist would, in fact, be beneficial. Nevertheless, in Section 6.4 we considered some grounds for treating this as a plausible hypothesis.

A further question remains, however. Under which circumstances do moral rights exist? This question is in one way more pressing for us than it would be for a traditional Rule Utilitarian, who seeks to explain rights in terms of ideal rather than actual compliance with or acceptance of rules. The idealizing nature of Rule Utilitarianism makes it well suited to explain the intuition that moral rights exist even when local practice does not seem to respect them. Are we forced to say, in contrast, that moral rights exist only when the local practice is to respect them?

Utilitarians are likely to want to steer a middle course here.⁴⁷ Unlike some rights theorists, they will not claim that all persons in all times and places have moral rights. They will deny that what G. K. Chesterton's Father Brown says about reason and justice applies to rights:

Reason and justice grip the remotest and the loneliest star. Look at those stars. Don't they look as if they were single diamonds and sapphires? Well, you can imagine any mad botany or geology you please. Think of forests of adamant with leaves of brilliants. Think the moon is a blue moon, a single elephantine sapphire. But don't fancy that all that frantic astronomy would make the smallest difference to the reason and justice of conduct. On plains of opal, under cliffs cut out of pearl, you would still find a notice-board, 'Thou shalt not steal'.⁴⁸

Utilitarians would not expect to find moral rights—or at least, not the same set of moral rights—in a world with plains of opal or cliffs cut out of pearl. According to utilitarians, moral rights ultimately have an instrumental justification of some sort, and so they are contingent on circumstances.⁴⁹ However, utilitarians will also want to avoid concluding that moral rights exist in only a very narrow range of circumstances, if they want to avoid being highly revisionary of our ordinary concept of moral rights.

According to the theory proposed in this chapter, some person *A*'s having a moral right consists in others' having reasons of the correct sort. To get a relevant case, let us suppose that *A* lives in a jurisdiction that recognizes no legal right not to be tortured, and that *B* faces a choice of whether to torture

⁴⁷ See Brandt 1992: 190–3.

⁴⁸ Chesterton 1911: 25–6. A talk by John Cottingham on 'Morality and the Transcendent' in September 2014 drew my attention to this passage.

⁴⁹ On the contingency of morality generally, see Mulgan 2011. Rawls 1999a: 137–9 claims that utilitarianism's conclusions, including about rights, are too dependent on 'general facts'.

A. Suppose finally that in this case *B* would produce slightly more utility by torturing *A* than by not doing so. The question is whether the theory proposed in this chapter can account for the strong intuition that it would be wrong for *B* to torture *A* in these circumstances, because doing so would violate *A*'s moral right not to be tortured.

To account for this intuition, we need to be able to claim that *B* has a strong reason not to torture *A* because this is his part in an actual pattern of respect for the right not to be tortured.⁵⁰ The actual pattern of respect cannot be one of legal recognition in *A*'s jurisdiction of a right not to be tortured—since, by hypothesis, that does not exist. However, this does not mean that there is no actual respect at all for the right not to be tortured. Others in *A*'s jurisdiction are likely to respect the moral right not to be tortured, in the sense that they refrain from acts of torture because they treat the fact that something would be an act of torture as a usually decisive reason not to do it. The same is true of others, further afield. Many of these people will respect the moral right not to be tortured not only in the strict sense of refraining from torture themselves because of it, but also in the broader sense of advocating that others, including others in *A*'s jurisdiction, also respect it.

Can we appeal to these instances of respect? Whether it is ad hoc to do so depends on whether it makes sense to think of *B*'s act of not torturing *A*, when combined with one or more of these instances, as forming a sufficiently beneficial and eligible pattern. That in turn depends in part on whether *B*'s act would be an irrelevant extension or elaboration of an existing pattern of respect. In the case of torture, it is hard to see why it would be an irrelevant extension or elaboration. *B*'s act, together with the existing instances of respect in his state and elsewhere, forms a pattern that realizes the benefits that ground the right not to be tortured—including, for example, the personal security of people in *B*'s state and elsewhere. As the number of people strongly disposed not to torture increases, we all gain a little more personal security, other things equal. Legal recognition of the moral right not to be tortured is helpful but not essential for a pattern of respect for this right to exist and be beneficial.

Of course, some moral rights are more closely intertwined with legal rights. We noted in Section 6.4 that some moral rights depend for their existence and content on legal rights. For example, my moral right to use your wheelbarrow depends on your legal property right in the wheelbarrow.

⁵⁰ More exactly: in a pattern that would be actual were *B* to play his part in it.

If the legal system in question is sufficiently unjust, your legal right may no longer generate my moral right. This means that some moral rights—those that depend on legal rights for their existence—simply do not exist in unjust circumstances. But this accords with common sense.

In sum, those moral rights that do not depend on legal rights will exist whenever there exists a pattern of respect for them to which the relevant agent could contribute. This pattern need not consist of legal recognition for the moral rights in the agent's jurisdiction. If, on the other hand, a moral right depends on a legal right, it will exist so long as the legal system in the jurisdiction is sufficiently just. Finally, in all of these circumstances there will be other kinds of reasons which are in some way associated with moral rights though not constitutive of them. These include act-based reasons not to torture, for example, that derive directly from the badness of the token act of torture, and reasons to advocate for respect for moral rights or for changes in the law.

6.6 Conclusion

Utilitarians have ample reason to try to develop a plausible theory of moral rights. Doing so would enable them to explain many common moral intuitions that are hard to explain otherwise. These include many intuitions about moral constraints, and especially about cases in which we believe not merely that an agent acted wrongly, but that she wronged someone. In this and other ways, the concept of moral rights is deeply entrenched in our moral and political thought. It could turn out that, according to the best theory, there are no moral rights, but that would be a surprising result.

The best prospects for a utilitarian theory of moral rights appear to lie with some kind of indirect theory, according to which the reason to respect a right on some particular occasion is not dependent on the value of conforming with it on that occasion. This may at first sound surprising, but it makes sense in light of the pluralistic theory of reasons we have been developing. The value of conforming with the right on a particular occasion instead explains an act-based reason—either an additional reason for conforming with the right, separate from the reason that constitutes the right itself or, if more could be achieved otherwise, a reason for infringing the right, to be set against the reason for respecting it.

We have been exploring the prospects for a particular kind of indirect theory, which seeks to explain reasons to respect rights in terms of actual

patterns of respect for them. The main challenges for this theory are, first, supporting the empirical claims it makes about the benefits of patterns of respect for rights and, second, explaining satisfactorily the conviction that moral rights exist even when they are not recognized by the relevant legal system. I have tried to give reasons for being optimistic on both counts. If these challenges cannot be met, utilitarians could instead explore further the idealizing theories of rights proposed by Rule Utilitarians.

Justice and Equality

This chapter discusses two distinct though related objections to utilitarianism: that it is inadequate as a theory of distributive justice, and that it is insufficiently egalitarian. These really are distinct objections, even though egalitarians naturally think that utilitarianism is inadequate as a theory of justice in part because it is insufficiently egalitarian. Some kinds of concern with equality do not fit within the theory of justice, and many non-egalitarians also claim that utilitarianism is an inadequate theory of justice.

Utilitarianism is in some important respects egalitarian, as I shall explain. One of the most significant of these respects is that it condemns as seriously wrong the inequality that marks the contemporary world. But it values equality in wealth or resources only instrumentally, and many egalitarians think that this misses some of the moral and political significance of this kind of equality. They object also to the specific distributive implications of utilitarianism. On the face of it, utilitarianism favours shifting resources from those who have expensive needs towards those who can extract more utility from them, who may be in every way better off. This sort of objection has seemed to many, especially in political philosophy, to be fatal to utilitarianism.¹ This is one reason why its stock is even lower among political philosophers than among moral philosophers.

When considered as a theory of justice utilitarianism is condemned even more widely. John Rawls and Robert Nozick—who in important ways have opposed views about justice—join forces in their criticism of utilitarianism as a theory of justice.² Participants in later debates amongst egalitarians

¹ Influential criticisms of the distributive implications of utilitarianism include Rawls 1999a: 19–30, Nozick 1974: 26–42, Sen 1979, and Dworkin 1978: 232–8 and 2000: ch. 1.

² Rawls's claim that utilitarianism 'does not take seriously the distinction between persons' 1999a: 24 is consciously echoed by Nozick's claim that using one person to benefit another 'does not sufficiently respect and take account of the fact that he is a separate person, that his is the only life he has' 1974: 30–1. Nozick goes on to accuse Rawls's theory of having the same flaw (228).

about the precise form of the best egalitarian theory were inclined to dismiss utilitarianism as not really even a candidate theory of justice.³

These critics had a point. It is not merely that each of them espouses a theory of justice that is at odds with utilitarianism, for the theories they espouse are also at odds with each other. The deeper point is that utilitarianism may seem not to have the conceptual resources to distinguish between justice on one hand and morality or goodness on the other. If the utilitarian theory of justice is simply that justice requires that social institutions maximize utility, for example, then its theory of justice coincides with its theory of the value of institutions. Nothing seems to be added by saying of the best institutions that they are also, for that reason, just. The complaint is then not so much that utilitarianism has the wrong theory of justice, but that it lacks a theory of the distinctive social virtue of justice at all.⁴

This chapter elaborates a more complex utilitarian theory of justice, according to which justice requires respect for moral rights. This enables us to distinguish justice from goodness, and so to explain the distinctiveness of justice as a virtue from a utilitarian point of view. Of course, the result is not a theory of justice that will satisfy all of those with egalitarian intuitions. But it will enable a more accurate estimation of the force of the challenges to utilitarianism from equality and from justice.

7.1 Distributive Justice

Our focus is on distributive justice in particular. This is a matter of justice in the distribution of benefits and burdens among people.⁵ This topic is broader

³ Amartya Sen's Tanner Lecture, delivered in 1979, discussed utilitarianism as a theory of equality: see Sen 1995. In an essay first published in 1981, Ronald Dworkin briefly did the same: see Dworkin 2000: 62–4. Both were of course critical of utilitarianism, claiming that it is inadequate as a theory of equality. A few years later, G. A. Cohen's survey of conceptions of distributive equality scarcely mentions utilitarianism: see Cohen 1989. Elizabeth Anderson's most influential contribution to the debate (1999) does not mention it at all, though in a later paper she mentions that Mill was a kind of egalitarian (2010: 3 n. 4). Mill himself noted that utilitarianism is often thought to provide an inadequate account of justice. He wrote: '[i]n all ages of speculation, one of the strongest obstacles to the reception of the doctrine that Utility or Happiness is the criterion of right and wrong, has been drawn from the idea of Justice' 1998/1861: 87. Economics-minded political philosophers are an exception to the trend of not discussing, or quickly dismissing, utilitarianism as a theory of justice. For example, see Roemer 1996: ch. 4 and Fleurbaey 2008: ch. 8.

⁴ John Tasioulas has claimed that the individualistic character of moral rights explains why justice poses a 'special challenge' for utilitarianism. See Tasioulas 2013: 772.

⁵ In Aristotle's formulation, it is concerned with 'the distribution of honour or money or such other assets as are divisible among the members of the community' 1976: bk. V, 1130b32.

than justice in the distribution of wealth and money. Among the relevant burdens and benefits are rights and liberties, powers, and social standing.

Even defined in this broad way, distributive justice is only one kind of justice. For tractability we will assume that it makes sense to discuss it separately from the other kinds, such as justice in war, or epistemic justice, or retributive justice. There are, no doubt, important connections between these topics, and a full treatment of any of them may have to discuss one or more of the others.⁶ But our aim is much more modest than this: we wish only to see whether utilitarianism can include a theory of distributive justice that is both distinct from its theory of the goodness of institutions and worth taking seriously.

One question is whether utilitarians should want to develop a theory of distributive justice at all. They could, instead, seek to answer questions about the proper distribution of benefits and burdens by appeal to their general theory of rightness and goodness. For example, they might say that an allocation of benefits and burdens is good insofar as it promotes well-being, and that an action which alters the distribution of benefits and burdens is right insofar as it makes the outcome as good as possible. Why should they want to say anything about justice, in addition?⁷

One answer is that convictions about distributive justice are an important subset of people's ethical convictions. Any ethical theory that seeks to explain people's convictions on the whole had better be able to explain convictions about justice, or to explain why they do not need explaining. Offering a theory of justice is one way to try to meet this demand. However, we might also wonder whether people are right to have convictions about justice. Do they play any important role not played by convictions about goodness or rightness?

There is no agreed or uncontroversial way of distinguishing justice from other normative concepts. Some try to distinguish it by saying that it applies to institutions rather than actions, though that runs the risk of construing it too narrowly.⁸ Others try to distinguish it by virtue of its alleged connections

⁶ For example, Matravers 2000 argues that political philosophers have been wrong to treat retributive and distributive justice as relatively separate topics, since both are about the moral justification of coercion.

⁷ See Brandt 1996: 199–202.

⁸ Rawls 1999a: 6–10 claims that the 'primary subject' of justice is the 'basic structure' of society, by which he means 'the way in which the major social institutions distribute fundamental rights and duties and determine the division of advantages from social cooperation' (6). Though Rawls allows that things other than the institutions of the basic structure may be just or unjust, he has been criticized for defining the subject matter of justice too narrowly: for example, see Okin 1989: ch. 5 and Cohen 1997. L. B. Murphy 1999 criticizes the idea that the principles governing institutional design are different from those governing personal conduct.

with the legitimacy of states or with rightful coercion.⁹ Others claim that justice is a matter of respect for people's rights.¹⁰ These views may be combined in various ways.

I will construe distributive justice as consisting in respect for people's rights in relation to the distribution of benefits and burdens. This fits with several commonplace claims about justice. For example, it fits with the idea that to have been treated unjustly is not merely for a wrong to have occurred, but to have been wronged. It also fits the idea that justice can be claimed or demanded, as of right. It fits the idea that justice is a matter of impartiality, since rights are also supposed to have that feature. It fits the idea that justice is an important part of morality, though not the whole of it. Finally, it fits the idea that justice is internally complex, in the sense that what justice requires depends on a balance of competing considerations—since what rights someone has also depends on a balance of competing considerations.¹¹

The downside of thinking of justice as consisting in respect for rights is that it risks construing justice too broadly. If Rita promised to lend Marcel her lawnmower, Marcel may have acquired a moral right to use it; but if Rita does not in the end lend it to Marcel, it may seem odd to say that Rita has behaved unjustly, or that Marcel has suffered an injustice.¹² One response to this would be to concede that the account is too broad, and to refine it so that it claims that justice is a matter of respect for moral rights *by institutions*. But this refined view seems too narrow: intuitively, individuals and their actions can also be unjust. This suggests a different response. If individuals and their actions can be unjust, perhaps we should accept that Rita's failure to lend Marcel her lawnmower was indeed unjust. The oddness involved in saying that it was unjust might reflect pragmatic considerations, deriving from the fact that it is not a very important injustice. We might be reticent to describe it as an instance of injustice for the same reason that we would be, in fact, reticent to describe it as having violated Marcel's moral right to use the

⁹ Buchanan 2003: ch. 5 claims that meeting a minimal standard of justice is a necessary condition of legitimacy. Rawls 1999b denies this. Rossi 2012 claims that the connection runs in the other direction: what justice requires depends on what is legitimate. Mill 1998/1861: ch. 5 claims that justice is concerned with those claims that society ought to enforce.

¹⁰ See, for example, Nozick 1974: ch. 7, Steiner 1994, and Tasioulas 2013.

¹¹ Raz 1984: 208–9.

¹² Lyons 1994: 84–5 claims that '[not] all violations of moral rights are injustices' (85), and gives as an example someone who has promised to go to a party: if he fails to go, Lyons says, this person has violated a moral right but he has not acted unjustly. I would treat this example in the same way as I treat the example of Rita and Marcel. I am grateful to an anonymous reviewer for pressing me on this point.

lawnmower. Both descriptions are overblown and inappropriate in normal circumstances, whether or not they are true.

If this reply fails to satisfy, we could hope to find some other restriction—short of the restriction to institutions—to distinguish injustice from other infringements of rights. Either way, we will proceed on the assumption that justice is either entirely a matter of respecting rights, or that respect for rights is a central requirement of justice. The distinctive ethical significance of justice will then be either wholly or partly explained by the distinctive ethical significance of rights. According to the view we have been developing, rights are one main source of constraints on the promotion of well-being. They are ethically significant in their contribution both to determining which actions are right and to the moral psychology of virtuous agents.

So we can distinguish distributive justice from other parts of morality, and utilitarians have some reasons to offer their own theory of it. If they do so they will be pitting their theory against a wide range of sophisticated contending views. These include egalitarian views of various sorts, according to which justice requires equality of well-being, or of opportunity for well-being, or of resources, or something similar to equality in one of these dimensions, such as sufficiency or priority for the worse-off.¹³ They also include the view that justice requires distribution of goods according to moral desert, so that it is unjust if sinners fare better than saints.¹⁴ Further, they include Nozick's historical entitlement theory, according to which justice places constraints on the way that goods are originally appropriated and on the ways in which goods are transferred, but not on the resulting patterns of distribution of goods.¹⁵ Utilitarians must hope to offer a theory of distributive justice that is, in comparison with these other views, sufficiently plausible to be worth taking seriously.

7.2 Justice for Utilitarians

Critics of utilitarianism typically assume that it must endorse one of two theories of distributive justice. The first of these is the simplest possible theory, according to which justice requires that goods are allocated in

¹³ See Sen 1995, Dworkin 2000: chs. 1–2, Arneson 1989 and 1999b, Raz 1986: ch. 9, Frankfurt 1987, Parfit 2000, Crisp 2006: ch. 6, Casal 2007, Otsuka and Voorhoeve 2009, and Shields 2012.

¹⁴ Aristotle 1976: bk. V, Ross 2002/1930: 35, 138–40, and Temkin 2000: 138–40.

¹⁵ Nozick 1974: ch. 7.

whichever way maximizes utility. According to this theory, 'justice' is a property of allocations of goods, and it is co-extensive with maximum goodness.

This theory is not at all plausible. It does not assign any distinctive role to the concept of justice, and its implications about justice are wildly out of step with ordinary convictions. Suppose, for example, that Ellen is about to donate her life savings to a very effective charity, when Joe comes along and steals the money and donates it to a slightly better charity, thereby making the outcome better. This theory implies that Joe's action made the world more just. This is, I assume, starkly at odds with ordinary intuitions about justice.

Utilitarians are committed to claiming that Joe's action made the world better. Depending on their theory of rightness, they may also claim that this makes Joe's action right. But it is very implausible for them to claim that, in addition, he made the world more just. This is to co-opt the concept of justice in a way that is both unconvincing as a claim about what justice consists in, and adds nothing worthwhile to the existing utilitarian claims about this case. We should reject this very simple theory.

In his discussion of utilitarianism as a theory of justice, John Rawls attributes a much more plausible claim to utilitarians. Rawls writes:

The main idea is that society is rightly ordered, and therefore just, when its major institutions are arranged so as to achieve the greatest net balance of satisfaction summed over all the individuals belonging to it.¹⁶

Generalizing slightly—and ignoring Rawls's methodological concern with only a single society—we can say that according to this utilitarian theory of justice, a society is just if and only if and because its 'major institutions' maximize utility. In contrast to the previous claim, this theory characterizes justice as a property of societies, which exists when their institutions are maximally good. Justice is then a matter of the goodness of institutions, not of allocations.

One advantage of this theory over the previous one is that it lets rights enter the picture. It is not certain, but quite plausible, that utility-maximizing institutions will assign rights to people, and it could be that these rights will match reasonably well our ordinary convictions about what justice requires. Rawls goes on, of course, to give reasons why he thinks this

¹⁶ Rawls 1999a: 20. Sartorius 1975: ch. 7 adopts a similar view, as does Shaw 1999: 217.

theory of justice will not be satisfactory all things considered.¹⁷ But whatever our view of the force of those criticisms, this theory is much more plausible than the simple theory we just considered. It may well be able to explain the conviction that Joe acts unjustly in stealing Ellen's life savings, for example, even though he makes the world better in doing so.

Whether it can do this depends on some details in the formulation of the theory, together with some hard to assess empirical claims about which institutions would maximize utility. The details of formulation concern how to spell-out the idea of 'best institutions'. One issue is the set of relevant alternatives—best out of which set of possible institutions?—and another is how to characterize the consequences of institutions. Depending on how we answer these questions, the theory Rawls attributes to utilitarians could end up at any point on a spectrum from recommending pragmatic piecemeal institutional reform, at one end, to a highly idealizing form of ideal theory, at the other.

In any case, this broad approach to thinking about justice is not, I believe, the way for utilitarians to go. It yokes justice too closely to goodness, and in doing so it suffers from a problem that also afflicts other theories of justice. The problem arises if we accept Rawls's idea of 'legitimate expectations', understood in a certain way. Legitimate expectations are entitlements generated by adherence to the rules of a sufficiently just society.¹⁸ For example, if you live in a sufficiently just society and you undertake the duties specified in your employment contract, you acquire an entitlement to the wages it specifies. This is not merely a legal right; if it were, there would be nothing unjust in changes to the law that cancel your legal entitlement. Your entitlement to your wages is also a moral right.

The crucial question is how just a society must be for legitimate expectations to generate moral rights in this way. Rawls does not answer this question. Most of his discussion of legitimate expectations occurs in the context of discussing ideally just societies, though at one point he makes remarks that suggest that he believes they generate moral rights also in somewhat unjust societies.¹⁹ In any case, I assume that legitimate

¹⁷ Rawls 1999a: 19–30, 137–53, 160–8, 284–5. These include its treatment of unjust preferences, the way in which the theory depends on facts, its alleged ignorance of the moral significance of the 'distinctness of persons', and the considerations Rawls gives for thinking that his principles would be chosen over utilitarian alternatives from the Original Position.

¹⁸ Rawls 1999a: 51–2, 74–6, 273–7.

¹⁹ He writes of cases in which the laws are unjust: 'How far we are justified in doing this [departing from existing norms], especially at the expense of expectations founded in good faith on current institutions, is one of the tangled questions of political justice. In general, all that can

expectations generate moral rights even in societies that are not perfectly just. If we do not assume this, we seem forced to say that arbitrary changes in the law which confound people's expectations are not unjust when they take place in imperfectly just societies.

The problem is what to say about property rights in sufficiently-though-not-perfectly just societies. We are assuming that members of these societies acquire moral rights to things in virtue of legitimate expectations. If we endorse a theory of justice which implies that, ideally, justice requires a certain allocation of property, or that it requires that a certain procedure is followed for acquiring property, we seem forced to allow that someone's property in some item could be both a requirement of justice (because he has a moral right to it generated by legitimate expectations) and disallowed by justice (because the ideally just allocation or procedure implies that it is someone else's property).

There are some ways of avoiding this problem. One is to deny the assumption we have been making, that legitimate expectations give rise to moral rights even in imperfectly just societies. That is a high price to pay, since we rely on this assumption in explaining the injustice of arbitrary changes to the law. A second way is to distinguish senses of 'required by justice', to avoid the threatened contradiction. We could say, for example, that in the imagined case justice 'non-ideally requires' that one person have the property, but 'ideally requires' that someone else have it.

This is, I think, the way that some theorists of justice hope to avoid this difficulty. But it is not very satisfactory, since it conflates two different issues. One issue is what justice requires in different circumstances: in the present circumstances, it requires that one person has a property right in this item; in other circumstances, it requires that someone else has a property right in it. A different issue is the nature of the requirement that justice imposes in any circumstance. The distinction between 'ideally requires' and 'non-ideally requires' on the face of it suggests that there is some defect in the way in which justice requires something when it non-ideally requires it.

Rather than distinguish ideal requirements from non-ideal requirements, we should simply distinguish what is required in ideal circumstances from what is required in other circumstances. In our imperfect society—provided

be said is that the strength of the claims of formal justice, of obedience to system, clearly depend [sic] upon the substantive justice of institutions and the possibilities of their reform' Rawls 1999a: 51–2.

it is sufficiently just for legitimate expectations to generate moral rights—justice requires that you get your wages.²⁰ This is not a second-rate requirement of justice, but a full-blown requirement of it. There is no superior way in which justice requires things, according to which it requires that someone else have this property. All that is true is that, in different circumstances, justice may require that someone else have it.

This suggests a way of thinking about justice which is highly procedural and which distinguishes much more sharply between justice and goodness than does the theory Rawls attributed to utilitarians. The rights that justice consists in respecting, on this view, are not those assigned by ideal or maximally good institutions. Instead, they are the moral rights that exist in the circumstances. In actual circumstances, they are the moral rights that people actually have. These include the core moral rights that are independent of any existing legal system, together with those moral rights which are generated by the mechanism of legitimate expectations, which depend on the features of the existing legal system. Someone's moral right not to be tortured is of the first kind, which implies that justice requires that she not be tortured regardless of whether her jurisdiction permits it. If she lives in a sufficiently just society—let us say, one in which most core moral rights are respected—then justice also requires that her legitimate expectations are satisfied. If, however, she lives in a grossly unjust society, the social rules will not generate moral rights through the mechanism of legitimate expectations. In that circumstance, justice requires only respect for the core moral rights—though prudence, efficiency, or other considerations may generate reasons for satisfying people's expectations, even though justice does not require it.

Utilitarians should, I believe, think about justice in this way. Perhaps surprisingly, it is a way of thinking according to which justice is significantly further removed from goodness than most contemporary theories of distributive justice would have it. But in identifying justice with respect for the moral rights that people in fact have, it assigns a distinctive and worthwhile role to the concept of justice. To act justly is to act within the constraints generated by people's moral rights. It will not always be possible to maximize the good and to act justly—as the example of Joe and Ellen illustrates.

This way of thinking of justice is in some ways similar to Nozick's theory. Like his theory, it defines distributive justice as simply a matter of respecting

²⁰ Here I assume our society meets this threshold, for the purpose of exposition. Whether it really does so is a separate matter.

people's moral rights, and not as involving the production of favoured outcomes.²¹ Also like his theory, it implies that any society in which all moral rights are respected is perfectly just. Justice is sharply distinct from other virtues, and a perfectly just society may be far from ideal in other ways. In contrast to Nozick, however, utilitarians will tend not to think of property rights as generated 'pre-institutionally', but instead as generated by people's behaviour against a background of sufficiently just institutions.²²

7.3 Kinds of Equality

In order to address egalitarian objections to utilitarianism we need to distinguish some different kinds of equality. This will also enable us to consider which, if any, of these kinds of equality will be requirements of justice according to the theory just outlined.

One kind of equality is what Rawls calls 'formal justice'.²³ This consists of applying rules or laws to people equally, in the specific sense that like cases are treated alike—where relevant differences are defined by the rules or laws themselves. This kind of equality is inconsistent with favouring some out of whim or personal preference, but it is consistent with favouring some because they are favoured by the law or rule itself. This kind of equality is not trivial or pointless: if the rules state that no-one may park in a certain area, and we have both parked there, but I get a ticket and you do not because you are friendly with the enforcement officer, we have been treated unequally in this sense. But neither is this sort of equality all that we care

²¹ See Nozick 1974: ch. 7. Rawls's theory is also procedural to a high degree: see Rawls 1999a: 73–8. According to the theory of moral rights proposed in Chapter 6, the general pattern of respect for moral rights promotes well-being. But it does not follow that each act of respecting moral rights promotes well-being, nor that perfectly just societies must contain maximally good institutions.

²² 'Pre-institutional' generation of property rights is generation prior to any system of positive law, as for example described by Locke 1988/1690: bk. 2 ch. 5. Utilitarians will tend to take the Hobbesian view (1946/1651: ch. 15) that property rights are 'post-institutional', in the sense that their existence depends on the existence of a system of positive law. As Murphy and Nagel note, this has important implications for the evaluation of tax regimes. They write: 'Private property is a legal convention, defined in part by the tax system; therefore, the tax system cannot be evaluated by looking at its impact on private property, conceived as something that has independent existence and validity. Taxes must be evaluated as part of the overall system of property rights that they help to create' 2002: 8.

²³ Rawls 1999a: 51. Rawls characterizes it as 'impartial and consistent administration of laws and institutions', and notes that it is compatible with slavery. See also Hart 1961: 153–9 and Hooker 2000: 46–7.

about. If the rules state that friends of enforcement officers may park where they like, I cannot complain on grounds of formal justice.

Utilitarians are likely to regard formal justice as important because of its instrumental value. Unless they are Rule Utilitarians and the rules in question are the rules in the ideal code, they will not regard equal and impartial application of rules as a fundamental moral requirement. Nevertheless, formal justice is constitutive of organizing things according to rules or laws at all: it is not an optional extra. Since there is often great value in organizing things that way, utilitarians will view formal justice as necessary to realize the instrumental value that is often achieved by organizing social life using rules and laws.

A second kind of equality is equal importance of equal benefits and harms. Utilitarianism is fundamentally committed to this, since it treats an equal-sized benefit or loss as making an equal-sized difference to overall goodness, no matter to whom the benefit or loss occurs. In the phrase that Mill used to sum up Bentham's view on this matter, 'everybody to count for one, nobody for more than one'.²⁴ This idea is entailed by the way that sum-ranking features in utilitarianism. As we have defined them, utilitarians claim that the value of an outcome is equal to the sum of the utilities it contains. This implies that an increment or decrement of one 'utile' makes the same difference to the goodness of an outcome, no matter to whom it falls.

Equal importance of equal benefits or harms is one way of conceiving the equal importance of persons. Sometimes it is said that it fetishizes utility, as if what it really expresses is a concern that *utiles* are treated equally. But this criticism is misplaced. The only reason that utilitarians care about *utiles* is that they are bits of goodness-for-people. The reason that they treat each *utile* as counting the same is that they treat people as equally important. On the other hand, it is also true that equal importance of equal benefits and harms is not the only way of conceiving the equal importance of persons. Prioritarians, for example, claim that benefits and harms matter more, morally, the worse off the subject. This might be thought to conflict with the idea that each person is equally important. But it is not clear whether it

²⁴ Mill 1998/1861: 105. Similar formulations can be found in Bentham's works. For example, he wrote: 'the happiness of the most helpless and indigent has as much title to regard at the hands of the legislator, as that of the most powerful and opulent' 1838–43d: 107. See also de Lazari-Radek and Singer 2014: 349 n. 19. For an interesting discussion of the respects in which Bentham's view may be more egalitarian than is usually assumed, see Postema 1998. For a similar argument in relation to Mill's views on justice, see Su 2013: ch. 4.

does so. Prioritarians do not attach extra importance to helping the specific individuals who are worse off, thereby treating those individuals as more important than others. Instead, they attach extra importance to helping the worse off, whoever they are. The extra importance goes with the level of well-being, not the individual. They can say to each person, 'we would have just the same concern for you, if you were as badly off'.²⁵ And, of course, egalitarians of different stripes offer their own conceptions of the equal importance of persons.²⁶

A third important kind of equality is equality of basic moral rights. Everyone who believes in moral rights at all agrees that different people have different sets of moral rights, since they can acquire or lose moral rights through their actions. Thus if I am rightly imprisoned, I have presumably lost, temporarily, some moral rights of free movement. If you keep your part of a just contract, you gain a moral right that others lack to the other party's performance. But an important and widely shared belief is that there is some set of *basic* moral rights that are equally held by all persons.²⁷

As we noted in Chapter 6, utilitarians who believe in moral rights are likely to think that what basic rights people have depends in some way—not necessarily in a highly sensitive way—on the circumstances. This follows from the fact that any utilitarian justification of moral rights will be, ultimately, an instrumental one. So utilitarians should accept that in at least some possible circumstances, people would not have moral rights, or not the ones we have. Therefore, utilitarians cannot support the claim that there is a set of basic rights that is possessed by all people in all circumstances.

However, they have some grounds for claiming that, in cases where some people have a specific set *X* of core moral rights, others in the same circumstances also have *X*. This is a narrower kind of equality of basic rights, since it

²⁵ Hooker 2000: 60–2 argues that this reply is insufficient, since we would not accept a parallel claim as proof of impartiality if it were made by an advocate of giving priority to the better off. Hooker may be right about this. However, I do not want to rely on that assumption here.

²⁶ Dworkin 2000: 62–4 and Kymlicka 2002: ch. 2 discuss utilitarianism as a conception of equality, and find it wanting. Dworkin 2000: pt. 1 presents his own theory of equality. Two other influential discussions are Cohen 1989 and Anderson 1999. The doctrine of the equal importance of equal benefits and harms expresses the sense in which utilitarians treat persons as equally important. However, it also extends to other entities capable of being harmed or benefited, and in so doing it conflicts with the common judgement that humans have superior moral status to other entities capable of being harmed or benefited, such as other animals. For different views on the deep puzzles involved in making sense of ordinary convictions about moral status see McMahan 1996, Kittay 2005, and Arneson 2015.

²⁷ This belief finds expression, for example, in adherence to the idea of human rights. See Tasioulas 2013. Buchanan 2003: 87–8 discusses utilitarianism and equality of human rights. On human rights in general, see Griffin 2008.

applies within circumstances, not across all possible circumstances.²⁸ Even this narrower kind of equality of basic rights is not guaranteed; it is not written into utilitarianism's fundamental principles. But it is supported by the same considerations that we appealed to in Chapter 6 in order to argue that the best patterns of respect for rights will not include making exceptions between cases. There we noted that the benefits of 'respect for rights' flow not merely from the behaviour such respect consists in, but also from the motivations it consists in. In particular, it seems likely that the motivation to depart from rules governing simple kinds of cases—such as the rule 'do not torture', governing opportunities to torture—will on the whole be counterproductive in utilitarian terms. If you believe that others are ready to break this rule when they judge that it would maximize utility to do so, your security will be undermined. The same seems to go for other basic moral rights, such as the right to freedom of thought or association.

If this is correct, it provides grounds for thinking that utilitarianism will support the narrower kind of equality of basic rights. The reason for this is that the possible grounds for beneficially distinguishing between persons also seem to be grounds for distinguishing between circumstances in respect of the same person. If we think it counterproductive for people to be motivated to treat person *S* differently just in case they judge her to have property *P*—where *P* corresponds with an opportunity to promote utility by departing from the usual rule governing cases of this kind—then we should also think it counterproductive for people to be motivated to distinguish between people according to whether they have property *P*. At least normally, the grounds for distinguishing between people will also be grounds for distinguishing between cases. Thus the reasons not to distinguish between cases will normally be reasons not to distinguish between persons.

There is, however, an important loophole. We are appealing here to hypotheses about the effects of different possible motivations. These effects operate through psychological and sociological mechanisms, not simply through chains of logical implication. If people conceive of others as divided into groups, they may be able, psychologically, to acknowledge different treatment of members of groups to which they take themselves not to belong, without suffering the harms of insecurity we have been appealing to. They may think, 'failure to respect *their* rights is compatible with secure respect for *mine*'. Unfortunately, we have ample evidence that people draw such

²⁸ 'Circumstances' are individuated by whichever facts ultimately explain the existence of a set of moral rights.

distinctions. For this reason, the utilitarian case for narrow equality of basic rights is incomplete. In conditions where social divisions are widely accepted, it is in principle possible for inequality of basic rights to be utility-maximizing.

This is an unwelcome implication of the utilitarian theory of rights. In this respect, it fails to capture some of our deepest convictions about rights and justice. Utilitarians can seek to mitigate the problem in at least two ways. One is to point out that where there are genuine grounds for distinguishing between groups—such as in the case of the groups ‘young children’ and ‘competent adults’—we are inclined on reflection to accept inequality of basic moral rights.²⁹ Our attachment to equal rights is not sweeping. The more problematic case is where the social division is based on mere prejudice. Utilitarians then seem forced to admit that, holding the prejudice as given, it is possible in principle that inequality of basic rights across the groups concerned would maximize utility.³⁰ However, they can emphasize that in cases such as this we should not hold the prejudice as given, and that it is very likely that there would be strong reasons for trying to reform or eradicate the prejudicial practice.

A fourth kind of equality concerns the material conditions of people’s lives in a way that goes beyond formal equality, equal importance of persons, or equality of basic moral rights. Egalitarians disagree amongst themselves over how exactly to specify the relevant concern here. Some claim that we should favour equality of welfare; others that we should favour equality of resources, or equality of opportunity for welfare, or equality in some other ‘currency’.³¹ They also disagree about whether it is, strictly, equality in one of these dimensions that we should favour, or instead some kind of sufficiency or some form of priority for the worse off.³² They disagree, thirdly, about the relevant sense or senses in which these things should be ‘favoured’. Some think that they are favoured in the sense of being required by justice, while others think that they are favoured in the sense of making the outcome better, and others think they are favoured in both senses.³³ But there is a broad sense in which they agree with each other that some form of what we

²⁹ On children’s rights see Freeman (ed.) 2014.

³⁰ See Rawls 1999a: 137–9 and Hooker 2000: 62.

³¹ Sen 1995, Dworkin 2000: chs. 1–2, Arneson 1989 and 1999b, and Cohen 1989.

³² Raz 1986: ch. 9, Frankfurt 1987, Parfit 2000, Crisp 2006: ch. 6, Casal 2007, Otsuka and Voorhoeve 2009, and Shields 2012.

³³ Parfit 2000: 84 distinguishes between ‘deontic’ and ‘telic’ forms of egalitarianism. For discussion of the usefulness of this distinction see O’Neill 2008. The distinction between ‘deontic’ and ‘telic’ views has been applied most often to forms of egalitarianism, but we can also draw it in relation to sufficiency and priority views.

might call ‘substantive equality’ is required by justice, and that utilitarianism goes wrong by failing to recognize this.³⁴

I have claimed that utilitarianism supports equality in the sense of formal justice, that it offers one conception of the equal importance of persons, and that with some important qualifications it supports equality of basic moral rights. The remaining question is thus whether utilitarianism is insufficiently egalitarian in the specific sense that it fails sufficiently to value substantive equality.

7.4 Utilitarianism and Substantive Equality

All utilitarians, as I have defined them, are committed to sum-ranking. This is the claim that the value of an outcome is equal to the sum of the goods (and bads) it contains. For utilitarians this means that the value of an outcome is equal to the sum of well-being it contains. This is the sense in which utilitarianism is, as is often said, ‘purely aggregative’. There is no room for manoeuvre over this, if we understand ‘utilitarianism’ in the way I have defined it.

However, consider the following example:

| | |
|----------|----------|
| (20, 40) | (30, 30) |
| World A | World B |

Each world contains just two individuals, whose well-being is represented by the numbers. In world A, the second individual is twice as well off as the first. In world B, both are equally well off, at a level midway between the two levels in world A. There are no further consequences associated with either world, so the numbers represent all of the well-being for each world, not just the well-being in it at some time or during some period. Utilitarians are committed to claiming that these two worlds are equally good.

Some take this claim to be sufficient reason to reject utilitarianism.³⁵ That is, they reject utilitarianism because it is incompatible with the judgement

³⁴ I borrow the term ‘substantive equality’ from McKerlie 1996: 274. So-called ‘relational egalitarians’ also believe that equality requires some form of substantive equality: see Anderson 2010.

³⁵ This example is modelled on one given in Roemer 1996: 160. He uses a more dramatic inequality to make the point, in which one person is 99 times as well off as the other. I have scaled it down to avoid worries about whether we can form reliable judgements about extreme cases. Williams 1973: 142–3 describes utilitarianism’s treatment of cases like these as ‘absurdly primitive’ and ‘just silly’.

that, in cases like this, the two worlds are not equally good. Since this judgement is incompatible with sum-ranking, endorsing the judgement gives us reason to reject sum-ranking—and, therefore, all forms of utilitarianism. Call this the *purely axiological* objection to sum-ranking. For it to be pure, we have to set aside any thoughts about possible causes or remedies of the inequality in World A—since these might activate intuitions about justice or duty. So we should imagine that the individuals in World A do not interact with each other in any way: let us say that each lives a lonely existence, separated by several centuries. To make it easier to grasp how their lives differ, suppose that they have the same, constant, level of well-being while alive, but that the first person lives for only twenty years, while the second lives for forty. Finally, for symmetry, make parallel assumptions about the individuals in World B.

With these clarifications in place, the purely axiological objection to sum-ranking does not seem to me to be very powerful. It is certainly not sufficient to defeat utilitarianism by itself. Utilitarians can agree that there is an important respect in which World B is better than World A: both individuals in B live better lives than the worse off individual in A. They can also agree that, if the worse off person in A could live for thirty years rather than twenty, that would be better for him. They merely insist that these facts are balanced by the respect in which A is better than B: the better off individual in A lives a better life than the individuals in B. It is not at all obvious that sum-ranking goes wrong in making this claim. Admittedly, sum-ranking makes the same claim no matter what the ratio between the levels of well-being of the individuals, so long as the total in each world is the same. As the ratio goes up, we may be more tempted to doubt whether sum-ranking could be right. But equally, as the ratio goes up, we should be more doubtful of whether we have grasped the features of the example accurately in forming our intuitions. The second doubt undermines the first. If the purely axiological objection works, it should work for the example specified, which we can readily understand.

The purely axiological objection would defeat *all* versions of utilitarianism were it to succeed against any, but it is too flat-footed to succeed. More interesting objections operate by appealing not simply to a verdict about a simple case, but also by offering an explanation of the relevant features of the case.

One such objection is that utilitarianism goes wrong by failing to recognize the noninstrumental value of substantive equality. As noted above, ‘substantive equality’ has to be specified further—in terms of well-being,

resources, opportunity for well-being, or some other currency. But faced with any objection of this sort, utilitarians will reply that the value of the specified sort of equality is purely instrumental, not noninstrumental. We will then have something of a standoff. In favour of the egalitarian side, a claim such as 'equality of well-being is good in itself' has some intuitive appeal, quite apart from its implications about cases. Against that, though, we have the powerful appeal of the levelling-down objection to this sort of claim.³⁶

The levelling-down objection has led some with broadly egalitarian sympathies to favour some form of sufficiency or priority instead.³⁷ In so doing, they side with utilitarians in claiming that equality itself has purely instrumental value. However, sufficiency and priority principles do not have as much intuitive appeal, independently of their implications about cases, as the egalitarian claims they displace. The claim that benefits matter more when they go to the worse off is appealing, I assume, primarily because it seems to explain egalitarian judgements about cases while avoiding the levelling-down objection. Taken just by itself it is certainly no more plausible than sum-ranking.³⁸

These points suggest that we should care more about whether a theory generates the right degree of support for substantive equality than about whether it endorses a claim about value that seems, in the abstract, intuitively appealing. Insofar as we think that substantive equality is valuable, it matters greatly whether utilitarians can explain this value. Their explanation will be in terms of its instrumental value, but for the time being we are setting aside intuitions about whether the value of equality is purely instrumental or also noninstrumental.

Considerations of diminishing marginal utility are sometimes used to claim that utilitarianism tends to favour one sort of substantive equality: equality of resources. Resources have 'diminishing marginal utility' in the sense that, usually, someone gains less well-being or utility from an additional unit of resources, the more resources she already has. This is an

³⁶ The levelling-down objection notes that, if equality is good in itself, it is in one way good if those who are better off are 'levelled down' to the position of the worse off, even if this would be better for no-one. See Parfit 2000: 98, 110–15. Temkin 2000 and McKerlie 1996: 286–96 discuss ways of responding to the levelling-down objection.

³⁷ For example, see Arneson 1999b: 497.

³⁸ See Hooker 2000: 59–65. As I noted above (n. 25), I do not want to take a stand here about whether priority principles are, all things considered, objectionably partial. I merely note that they are not especially plausible in the abstract, independently of their implications about cases.

empirical claim with some exceptions, but it is nevertheless an important and credible generalization.³⁹

A very simple argument that utilitarianism supports equality of resources states that, given this generalization, we can typically increase overall well-being by equalizing resources, since the loss to the rich will be outweighed by the gains to the poor. This argument sets aside the effects of inequality on productive output, and asks instead how to distribute a fixed amount of resources so as to maximize utility. Even with this caveat, however, this argument fails.

The reason the simple argument fails is that the generalization that, usually, *each* person will gain less from each additional unit of resources does not imply that *different* people make gains at the same rate. Nor does it seem to be true that different people make gains at the same rate. Some people, for example, have more expensive goals, ambitions, or tastes than others. Some are also inclined to expect more of their goals, ambitions, and tastes to be fulfilled than are others. Some are more adept at pursuing their goals than others. Furthermore, some people have expensive medical or other needs not shared by others. All of these facts make it likely that different people gain personal utility at different rates—all of which is consistent with the generalization that, for each person, additional units of resources typically bring smaller gains in utility.⁴⁰

In a fixed population, total utility is maximized when marginal utility is equalized. If different people are on different ‘utility curves’ (that is, the function from resources to utility is different for each person) their marginal utilities will be equalized at different levels of resources and of utility. Consider Figure 7.1. The marginal utilities of Bob and Sheila are equalized, and so their joint utility is maximized, when they are placed at points on their respective utility curves that have equal gradients. As Figure 7.1 shows, this leaves Bob with more resources and more utility than Sheila. The reason for this is that he gains more utility, in general, than Sheila does from each unit of resources—his curve is steeper than hers, at each level of resources. Redistributing resources from Bob to Sheila to equalize their amounts of resources would decrease total utility, since she would gain less utility than he would lose.

³⁹ Sometimes this assumption is defended using an a priori argument: see, for example, Lerner 1944: 26–7. Such arguments are not very persuasive however—if only because there are some exceptions, as when an additional item will complete someone’s collection, or take her over some other threshold. See Frankfurt 1987: 24–30 and Shaw 1993: 171–2 for discussion.

⁴⁰ See Shaw 1993: 179–81.

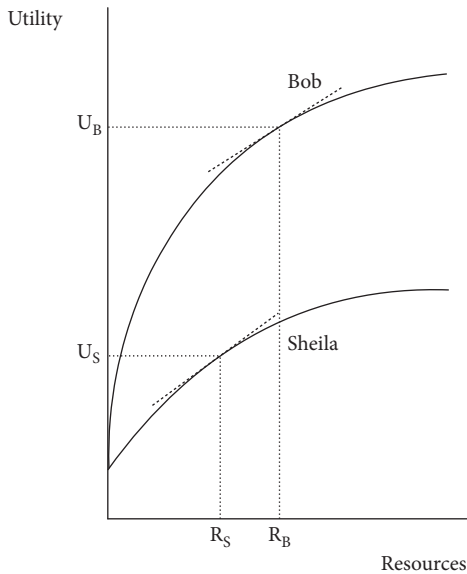


Figure 7.1 Why the simple argument does not work

This means that an omniscient utilitarian administrator concerned to maximize the utility gained from a fixed amount of resources will, in allocating them so as to equalize marginal utilities, almost certainly give different people different amounts of resources, thereby leaving them with different amounts of utility. If substantive equality means equality of welfare or equality of resources, this administrator would not produce substantive equality.

However, we should not think of substantive equality in this way. We do not choose directly between allocations of resources, but instead between institutions or general policies that have somewhat predictable effects on allocations. Second, we should not ignore the effects of distributive policies and institutions on production, or the total stock of goods. Since what matters for utilitarians is the whole future, they will also want to bear in mind issues of sustainability rather than aiming for maximum utility for currently existing people.⁴¹ Still, we do not accurately model that problem by

⁴¹ John Stuart Mill's remarks on the stationary state of the economy are interesting in this connection. He pointed out that continued economic growth may lead to overpopulation and ecological damage: 'It is not good for man to be kept perforce at all times in the presence of his species... Nor is there much satisfaction in contemplating the world with nothing left to the

assuming a fixed stock of goods. Third, and perhaps most importantly, there are no omniscient utilitarian administrators. I argued in Chapter 4 that it is very difficult even for an individual to know what would be best for her. The idea that any administrator could know how to maximize utility for a population of people is so unrealistic as to be not even a useful simplifying assumption when considering these matters.

The economist Abba Lerner argued in 1944 that lack of knowledge provides a stronger utilitarian argument for equalizing resources (again, setting aside the effects of distribution on production). He pointed out that, if we do not know the utility curves of different individuals, and we assume that the marginal utility of resources is diminishing for each person and that utility curves are randomly distributed, then we maximize expected utility by equalizing resources. The reason for this is that any departure from equality would be just as likely to increase as to decrease utility, but because the marginal utility of resources is diminishing for everyone, increases would on average be smaller than decreases. Under these conditions, *expected* utility would be maximized by equalizing resources.⁴²

Lerner's argument is ingenious, but it is limited in its practical significance. One reason for this is that, as he emphasizes, the argument assumes a fixed stock of goods, and thereby ignores the effects of distribution on production. Another is that the argument assumes that the distribution of utility functions across the population is random. This assumption is necessary to support the claim that departures from equality are equally likely to decrease actual utility as to increase it.⁴³ A third reason is that we are not as ignorant about differences in utility functions as Lerner's argument assumes. We know, for example, that some people have very expensive and persistent medical needs, and it is very plausible that they require more resources than

spontaneous activity of nature; with every rood of land brought into cultivation, which is capable of growing food for human beings; every flowery waste or natural pasture ploughed up ...' 2008/1848: 756.

⁴² See Lerner 1944: ch. 3 and Lerner 1978. See also Brandt 1979: 313–16 and Barry 1989: 176–8.

⁴³ Pogge 1995 emphasizes the importance of this assumption, and claims that Lerner is not entitled to it. It may be that Lerner lacks sufficient grounds for the assumption—but Pogge also makes the stronger claim (337) that it is inconsistent with Lerner's starting assumption that we do not know individuals' utility functions. This stronger claim seems to me to depend on conflating ignorance about individuals' utility functions with ignorance about the distribution of individuals' utility functions. In any case, Pogge is surely right to highlight the importance to Lerner's argument of the assumption that utility functions are distributed randomly.

others to reach any given level of well-being. This is true, for example, of people who require expensive ongoing medical or social care.⁴⁴

The upshot is that, in considering the extent to which utilitarians can support substantive equality, we should neither assume omniscience nor exaggerate ignorance. We know some things about the ways in which distribution of resources affects overall utility, and some of these help to make a utilitarian case for substantive equality.⁴⁵ We know much less about how general policies would affect specific individuals—though even here we may know, for example, that more or less generous state-funded health and social care would help or hinder individuals with specific kinds of needs. We also know that what maximizes the utility of currently existing people may well not maximize utility over the long run, taking future people into account.

In light of these considerations, we can tentatively conclude that there is some tendency for utilitarianism to favour equalizing resources, but that this has to be set against considerations of productive growth and sustainability on one hand, and known differences in individual needs on the other. Since egalitarians also typically seek to balance equality against considerations of productive growth and sustainability, the key issue in considering whether utilitarianism sufficiently values substantive equality is its treatment of known differences in individual needs.

7.5 Known Expensive Needs

Where an individual is known to have more expensive needs than average, an opportunity exists to promote utility by shifting resources *away* from her. But doing this would leave her with fewer resources and less well-being than others. This is the most pointed egalitarian objection to utilitarianism.⁴⁶

⁴⁴ Lerner 1944: 40 recognizes this.

⁴⁵ Examples of relevant empirical research include Hsieh and Pugh 1993, Lorant et al. 2003, and Lynch et al. 1998. Ott 2005 presents evidence from seventy-eight countries that average happiness is positively correlated with equality of happiness. O'Neill 2008 is a useful philosophical discussion of the likely instrumental value of equality. Many of the benefits O'Neill identifies (for example, fraternal social relations, and non-domination) seem likely to promote utility. Mary Wollstonecraft may also be right that 'virtue can only flourish amongst equals' 1790: 141.

⁴⁶ For an influential statement of the objection, see Sen 1995: 313–15. For a discussion that is sympathetic to utilitarianism see Stein 2006.

To get the objection in clear focus we must remember a couple of important points. One is that utility-maximizing institutions and policies will not in every case lead to utility-maximizing actions or allocations of goods. There are costs to tracking and responding to the differences between cases, and these will often be greater than the benefits. These costs are partly administrative, but they are also a matter of security, as we have already mentioned. Further, institutions or policies that grant the discretion needed to take account of individual differences may concentrate power in ways that are counterproductive from a utilitarian point of view. For these reasons, the best institutions and policies will typically apply general rules that do not track every knowable difference between people.

Second, the really crucial issue is not, in fact, whether someone has expensive needs, but whether her marginal rate of utility is lower than average. As Mark Stein has argued, discussions of disability often wrongly equate having a low level of well-being with having a low marginal rate of utility. Someone with mobility problems might have a lower level of well-being than someone without those problems, if they have the same amount of resources. But it does not follow that she also has a lower marginal rate of utility. In fact, her marginal rate of utility may well be higher than the other person's—perhaps because the next increment of resources would purchase a wheelchair that greatly improves her mobility. For this reason, there will often be strong utilitarian grounds for allocating more resources to those with expensive needs. As this shows, we need to distinguish two senses of 'expensive need'. In one sense, someone's need is expensive just in case it costs more to meet this person's need than it does to meet the same need for others. In a different sense, a need is expensive when meeting it brings about less additional well-being than could be achieved with a different use of the same resources. The crucial point is that the first sense of expensive need does not entail the second.⁴⁷

The relevant case, then, is one in which someone has an expensive need which is trackable by institutions and policies and which is such that she has a low marginal rate of utility. One context in which this sort of case arises is medical treatment of people near the end of their lives. Such people may be young or old, and they may have had very good or much less good lives. Some will think that—for broadly egalitarian reasons—we should spend more to give a young person who has not had a good life an extra month

⁴⁷ Stein 2006: 41–5. Failure to note this is a fallacy of much discussion of disability. However, Sen 1995: 313 is careful to avoid it.

of life, than we should spend to give an extra month (at the same level of well-being) to an older person who has had a good life. Or consider a second case, in which we have two patients who have had equally good lives so far, but we can extend the life of only one of them. One of them is now depressed and expected to remain so, while the other is not. The depressed person appears to have a lower marginal rate of utility, since the life-extending treatment would add less to his well-being than it would add to the other patient's well-being.

These are extremely difficult decisions, and we should be appropriately tentative in our judgements about them. In the case of the young and the old person it is hard not to assume that the extra month of life would add more if given to the young person than it would if given to the older person. But if we accept the stipulation that the extra well-being would be the same either way, it is not clear, to me at least, that we should favour the younger person. Effects on third parties may of course tilt the balance, but focusing just on the benefits to the patients, I have no firm conviction that we should favour either of them.

In the case of the depressed person, it seems to me that we have broader reasons for maintaining a policy of non-discrimination. The striking thing is that we seem to accept some differences in expected quality of life as grounds for allocating medical resources, but not others. If *physiological* differences meant that some scarce drug would completely cure patient *A*, but only partially cure *B*, and we could treat only one of them, we would almost certainly favour treating *A* over treating *B*. But if the reason for the difference in expected quality of life relates to certain other characteristics of persons—including mental health, gender, race, or sexuality—we would not treat the difference as a reason to favour one person over another.⁴⁸ What explains this complex combination of attitudes?

The combination is hard to explain on purely egalitarian grounds. What needs to be explained is responsiveness to differential benefit in some cases but not in others. If we start from an egalitarian principle, the default policy would presumably be not to respond to differential benefit. That gives us the right answer for the cases in which we think differential benefit is not a ground for favouring anyone—for example, where it is due to depression or differences of race or gender. But the question then is why we should favour on grounds of differential benefit on other occasions—for example, where it

⁴⁸ See Stein 2006: ch. 11 and Franklin 2017.

is due to purely physiological differences. It cannot be the size of the benefit that explains this, since that is not reflected in the structure of our attitudes: we do not think that differences in benefit due to gender or race become eligible grounds when they reach a certain magnitude.

Our complex attitudes here seem more easily explicable if we start from a utilitarian position. Utilitarianism suggests that the default policy is to respond to differential benefit. That gives the right answer for the cases in which differential benefit has purely physiological causes. The question is then how to explain why we should not favour on grounds of differential benefit when it is due to characteristics such as mental health, gender, race, or sexuality. But that may well be explicable as a utility-promoting anti-discrimination policy. The policy of treating some differences as irrelevant to these medical decisions may be justified, on utilitarian grounds, because it promotes general security, encourages social solidarity, and removes disincentives for people to be open and unashamed about these central features of their identity.

In this way, it may be easier to capture the complexity of our considered judgements if we start from utilitarian rather than from egalitarian principles. In many cases we do not know enough about individuals to allocate resources so as to maximize well-being. Because we are ignorant, the best we can do may be to adopt policies and construct institutions that treat people equally. In other cases we do know about relevant differences between persons, but once we take the costs of tracking differences into account (that is, administrative costs and costs in terms of people's security), there remain utilitarian reasons for treating people equally.

In the most interesting sort of case, however, we know about differences—and the costs of tracking them are outweighed by the potential benefits of doing so. I claimed that we are likely to have a complex set of attitudes to these cases. With respect to some of them, we will count differential benefit as a reason for allocating scarce resources to one person rather than another, in line with a straightforward application of utilitarianism. With respect to others, we will think it wrong to count differential benefit as a reason to allocate to one person rather than another, and endorse a form of equal provision instead. I claimed that utilitarianism may be best placed to explain this combination of attitudes. Our commitment to equal treatment may be a specific anti-discrimination policy, justified on more indirect utilitarian grounds. If that is correct, utilitarianism may turn out to be the best theory of our commitments to substantive equality.

7.6 Conclusion

Utilitarianism is capable of distinguishing between justice and goodness. On the view outlined in this chapter, distributive justice is a matter of respect for persons' moral rights. Individuals, actions, policies, institutions, and societies can all be just or unjust, insofar as they can either respect or fail to respect moral rights.

The utilitarian theory of distributive justice accommodates several kinds of equality. As a matter of basic principle it is committed to equal importance of equal benefits and harms. Further, it tends to support equality in the sense of formal justice, and it also tends to support—at least in good circumstances—what I called narrow equality of basic moral rights.

I also claimed that, in sufficiently just societies, persons' moral rights depend in part on their legal rights, including their legal property rights. If this is correct, it suggests that utilitarians should treat substantive equality as a goal of good policies rather than a requirement of justice. It cannot be a requirement of justice according to the theory I have outlined. If it were, it would be possible for justice to entitle two different people to the same property at the same time—one as a result of legitimate expectations, and the other as a result of the demands of substantive equality. Justice would then make incoherent demands. Instead, we should think of substantive equality as a goal to be pursued, while recognizing that the requirements of justice can conflict with it.

This raises the question of the relative importance of justice and substantive equality. The importance of justice is determined by the importance of moral rights. In Chapter 6 we treated these as usually but not necessarily decisive in determining the rightness of actions. For example, the constraints that rights place on action explain why utilitarianism need not be committed to imposing great harms on some for the sake of small benefits to many others. On the other hand, the constraints placed by rights are not absolute: sometimes it is morally right to infringe a right. It follows that it is usually but not always right to be just. Justice is important, but it is not lexically prior to other considerations.⁴⁹ I take this to be a welcome implication of the theory. The fact that we believe that justice should sometimes be tempered with mercy, for example, indicates that we do not think that it is lexically prior to all other considerations.

⁴⁹ Contrary to what Rawls 1999a: 3 claims.

Virtuous agents and good governments will be disposed not to act in ways they consider to be unjust without what they take to be very strong reasons. In this psychological sense, too, justice places a strong constraint on our actions. Despite this, it will sometimes be virtuous, or good policy, to pursue other goals, including equality, even when doing so conflicts with the requirements of justice. Recognition of the messy trade-offs that this involves is a strength of the utilitarian theory.⁵⁰

⁵⁰ Sidgwick 1907: 444 wrote: 'the Utilitarian view that the disappointment of natural expectations is an evil, but an evil which must sometimes be incurred for the sake of a greater good, is that to which Common Sense is practically forced, though it is difficult to reconcile it with the theoretical absoluteness of Justice in the Intuitional view of Morality.'

8

Legitimacy and Democracy

According to the simplest forms of utilitarianism, nothing much distinguishes politics from other areas of human activity. These forms picture politics as an arena in which the promotion of utility is the only relevant normative consideration, just as it is elsewhere. The only contrasts between attempting to promote utility through political action, and attempting to promote it in other ways, lie in the scope of decisions and the means available.

We have been developing a somewhat more complex form of utilitarianism, which recognizes reasons to participate in worthwhile patterns of action as well as reasons to promote utility directly. I claimed that this enables utilitarianism to explain moral rights, and consequently to give a plausible account of justice as respect for moral rights. I also claimed that utilitarians have significant reasons to pursue substantive equality as a policy aim. These features of the view developed in this book give it a somewhat more complex picture of the considerations that should govern political decisions.

Yet this picture, so far, still portrays politics as governed by just the same considerations as other parts of life. In our non-political lives we should also recognize reasons to participate in good patterns of action; we should respect people's moral rights and at least not frustrate the aim of substantive equality. The extra complexity we have added to utilitarianism, so far, has not yielded any grounds for treating politics as importantly different from other realms of activity. As with the simplest picture, the distinctiveness of politics is only a matter of the means available to political agents and the scope of the decisions they face.

The idea that politics is in this way continuous with other parts of life runs counter to many arguments in contemporary political philosophy. These arguments emphasize the distinctiveness of politics, and claim that it is a mistake to think of political philosophy as, simply, the application of moral principles to political decisions. Strikingly, Rawlsians and 'political realists' converge on this conclusion, even though they argue from quite different premises.¹

¹ Rawls expresses the distinctiveness of politics through his ideas of public reason and of a political conception of justice. See Rawls 1996 and 1997. 'Political realist' is a label applied to a

There seems to be something right about these claims that politics is distinctive. One way to think of them is as articulating a kind of restraint that, we might think, is appropriate when citizens address each other.² We know that there is deep disagreement about moral matters, but also that we must get on with the business of finding ways to live together. In this circumstance, it seems inappropriate to conduct discussions about political issues as if they were simply a matter of each citizen or political leader applying her own favoured moral theory, or trying to persuade others of its correctness. To do this is to propose to conduct political life on terms one does not expect other reasonable people to accept. Persistent disagreement seems to have different significance in politics than it has in a moral philosophy seminar, even though it is not easy to spell out precisely what the difference is.

Utilitarians are not likely to end up agreeing with either Rawls or political realists about the considerations that should govern political action. However, they may still have something to learn from the emphasis on the distinctiveness of politics that is characteristic of contemporary political philosophy. In this chapter I shall argue that utilitarians can and should take account of the importance of legitimacy in politics. Doing so enriches their account of politics and bolsters their support for democracy. Whether a decision or procedure is legitimate often makes a difference to what should be done. Though the principles that apply to political decisions are, in the end, the same as those that apply to other decisions, politics is an arena in which considerations of legitimacy are especially likely to be important.

8.1 Government House Utilitarianism

Two features of utilitarianism make it especially likely to result in a technocratic model of politics. One is its monistic theory of value. Ethical views that recognize more than one fundamental value typically allow that, within

somewhat diverse collection of political theorists who claim that it is a mistake to think of politics as applied moral philosophy. Williams 2005: ch. 1 is an influential statement, and Galston 2010 is an influential survey.

² This is explicit in Rawls: 'Citizens realize that they cannot reach agreement or even approach mutual understanding on the basis of their irreconcilable comprehensive doctrines. In view of this, they need to consider what kinds of reasons they may reasonably give one another when fundamental political questions are at stake' 2007: 766. See also Rawls 1999a: 311–12 on the 'duty of civility'.

broad limits, it is not clear how the balance between competing values ought to be struck. One possible role for political discussion and debate is then to decide exactly how to strike such balances. Because it recognizes only one kind of fundamental value, utilitarianism cannot assign that role to political discussion.

The second feature is that utilitarianism places so much weight on facts. This is especially true of Act Utilitarianism, but it is true to a lesser degree of other forms as well. Utilitarians treat what for others are matters of principle as instead contingent on the facts of the case. Utilitarians cannot reason directly from principles to practical conclusions, but almost always have to take account of the circumstances at hand. In particular, of course, utilitarian conclusions depend on facts about what would promote people's well-being. Utilitarianism has, as we might put it, a thirst for information.

Taken together, these two features encourage a technocratic model of political decision-making. They encourage a model according to which the end of political action is settled, leaving only questions of means to be decided. And since the appropriateness of means depends so heavily, according to utilitarians, on complex information about people's tastes and circumstances and on complex truths about the operation of economic and social institutions, it is natural to conclude that it should be left, for the most part, to experts. Politics is thus the application of moral principles in the light of expert judgement about matters of fact.

It is no surprise that this tendency of utilitarianism to support a technocratic model of politics has been noticed both by utilitarians themselves and by their critics. Early in his career, Bentham thought that enlightened despotism might be the best way to organize politics for utilitarian purposes.³ More recently, Robert Goodin has embraced a broadly technocratic model of politics.⁴ Among the critics, Rawls has remarked on the tendency of utilitarianism to think of political issues as a matter of 'higher order administrative decision'.⁵

The administrative model of politics—of politics as the pursuit of a given end by expert judgement of appropriate means—has something to recommend it. Some political decision-making surely is like that, at least in its dependence on expert judgements about matters of fact. On the other hand,

³ On Bentham's 'conversion' to democracy see Schofield 2006: chs. 4–5.

⁴ See Goodin 1995: chs. 1 and 4.

⁵ Rawls 1999d/1958: 65. He is here describing the way he takes utilitarians to think of justice, but the point presumably applies more generally.

much political discussion also concerns which ends are appropriate. For all that we have said so far, utilitarians seem bound to think of this discussion of ends as ultimately regrettable—a practical obstruction to the business of administering things to the greatest public good.

The technocratic model is, in one way, 'elitist'. Since it portrays political decisions as, properly understood, ultimately matters of complex judgements about facts, it strongly suggests that political decisions should be left in the hands of experts—technocrats or 'epistocrats'. This is the sense in which John Stuart Mill is often said to have been elitist, when he advocated plural votes for the well-educated.⁶ On the other hand, this kind of elitism is not obviously inappropriate if political decisions really do boil down to complex judgements of matters of fact.⁷

Other elements of the utilitarian tradition are elitist in a different, and more objectionable, sense. According to some indirect forms of utilitarianism, the rules of ordinary morality have a certain sort of utilitarian justification as useful heuristics or decision rules. The idea is that it is better in most cases to decide what to do by applying these rules than by attempting to make a utility calculation. However, it might be bad on utilitarian grounds to encourage people to think of these rules as mere heuristics, for that might encourage them to look for the exceptions, which might do more harm than good. Thus, as Sidgwick noted, it might be better on utilitarian grounds to keep secret the underlying utilitarian rationale for common moral rules; moreover, the utility of keeping this secret should perhaps itself be kept secret.⁸ Sidgwick went on:

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.⁹

In these passages Sidgwick is joining together two ideas: the idea that common moral rules might be a utility-promoting heuristic or decision procedure, and the idea that it would promote utility to keep this rationale for using the common moral rules secret from 'the vulgar'. This

⁶ See Mill 1991c/1861: ch. 8.

⁷ See Estlund 2008: ch. 11.

⁸ He wrote: 'it seems expedient that the doctrine that esoteric morality is expedient should itself be kept esoteric' Sidgwick 1907: 490. His striking turn of phrase was somewhat self-defeating: this is one of Sidgwick's best-known claims.

⁹ Sidgwick 1907: 490.

combination of ideas appears to be what Bernard Williams meant by 'Government House utilitarianism', or what he described as 'the social embodiment of indirect utilitarianism'.¹⁰

This combination of ideas is certainly troubling.¹¹ It expresses a more objectionable kind of elitism, and it carries the colonial overtones alluded to by Williams's choice of name for the view. If we view it as a model for politics, it may suggest that utilitarians should favour 'enlightened' elites making decisions on secret grounds for the sake of the 'vulgar' masses.

However, we should note that the idea that common moral ideas are useful heuristics does not mandate the social division of labour and knowledge that Sidgwick entertained. We could instead think that it is a responsibility of each virtuous agent to evaluate her own decision procedures, so that the division between use of heuristics, and evaluation of heuristics, occurs within rather than between persons.¹² Those sympathetic to Sidgwick might point out that it is hard to deny that people have different cognitive abilities, and so it is hard to deny that they could be expected to be more or less able to perform the calculations necessary to evaluate ordinary moral ideas. That is true: but it need not follow that a social division of labour in evaluating moral codes would promote utility. If we are all roughly equally bad at evaluating the consequences of belief in moral codes, as is somewhat plausible, the differences between us may not matter much.

We have identified two possible utilitarian models of technocratic rule. In the first, experts make political decisions because they are most able to make the complex judgements required to identify the best means of promoting utility. The second model adds that the utilitarian grounds of these decisions should be kept secret from ordinary people, for the reasons that Sidgwick explained. An important question is whether one or other of these models must be the right way for utilitarians to think of politics.

¹⁰ These phrases are from Sen and Williams 1982b: 16. Williams's critique of Government House utilitarianism is expressed briefly there and at slightly greater length in Williams 1973: 135–50 and Williams 1985: 108–10. It is also touched upon in Williams 2005: ch. 4. See also Rawls's discussion of publicity: 1999a: 115, 397–8.

¹¹ As Sidgwick himself noted: he referred to it as a 'swarm of perplexities and paradoxes', but thought that it might be 'a necessary evil' 1907: 490.

¹² This seems to be what Brink and Railton have in mind, for example. See Railton 1984: 151–6 and Brink 1986: 426 n. 23 and 427–30. Williams 1985: 109 objects that this sort of view relocates the damaging division to each individual's psychology, claiming that it requires 'some kind of willed forgetting'.

8.2 Democracy as Eliciting and Aggregating Preferences

It is possible to find a role for democracy within these technocratic models. In fact, one of the traditional utilitarian arguments for democracy fits neatly into them. This is the argument according to which democratic processes are justified because they elicit and aggregate information about people's preferences.

In order to know how to promote utility, governments need to know what would be good for people. It is hard to deny that people's preferences are one relevant source of information. According to what is known as the 'aggregative conception' of democracy, elections have an important role to play in generating knowledge of what would promote the good. For example, Jonathan Riley has argued that, when gathering cardinal and interpersonally comparable utility information is impossible or too costly, utilitarians ought to be democrats. He writes:

Democratic government is seen as a set of procedures for calculating the general welfare in environments where a full-blown utilitarian calculus is too costly to establish and run. In particular, democratic decision making takes account of the fact that personal utility information is initially dispersed and costly to retrieve, compare, and aggregate.¹³

On this view, elections can perform the valuable epistemic functions of getting people to express their preferences and aggregating those preferences into a social choice.

There are many well-known limitations to these functions of democracy. For one thing, the relationship between someone's preferences as expressed in an election, and her well-being, is likely to be quite distant. Her preferences might express her overall attitude to the political issue under consideration, taking into account considerations other than her own well-being. If they are narrowly focused on her own well-being, they may nevertheless not be a reliable guide to what would be good for her. Though there is no doubt some truth in the saying that each person is the best judge of her own interests, it is a hedged and qualified truth.¹⁴ A person's preferences may reflect inaccurate beliefs, or for other reasons fail to identify what would be

¹³ Riley 1990: 343.

¹⁴ Mill wrote: 'If a person possesses any tolerable amount of common sense and experience, his own mode of laying out his existence is best, not because it is the best in itself, but because it is his own mode' 1991b/1859: 75. This is of course compatible with recognizing the likelihood of error in prudential judgements.

good for her.¹⁵ Interestingly, there is also some psychological evidence that, in one important respect, people make better decisions about future benefits for strangers than about future benefits for themselves or those close to them.¹⁶

Second, there are well-known problems with the idea that elections aggregate preferences. Social choice theory has generated many proofs—of which Arrow’s theorem is the best-known—that no aggregation procedure can satisfy small sets of plausible requirements. Of course, there are debates about the normative significance of the requirements these proofs assume, and about the relevance of the formal results to real-world politics.¹⁷ But without entering those debates we can at least say that the problems identified by social choice theory make it difficult to spell out coherently the idea that democracy aggregates preferences.

In any case, the epistemic functions of democracy in eliciting and aggregating preferences are limited in other, very straightforward, ways. One is that the electorate is not coextensive with the set of persons or other entities with interests at stake in political decisions. Voters are not disqualified from voting on matters in which their interests are not at stake; more significantly, many whose interests are at stake are not voters. This group includes children, citizens of other states, non-human animals, and—perhaps most importantly of all—future people.¹⁸ A second and interlocking problem is that, if the purpose of democracy is to elicit and aggregate information about voters’ preferences, it is not clear why utilitarians should prefer democracy to voter opinion research. Such research could survey the opinions of a representative sample of voters in more depth and at much less cost than is achieved in an election.¹⁹ As these two problems suggest, it is not at all clear whether standard democratic assumptions about the extent of the franchise

¹⁵ Some reasons for this were emphasized in Chapter 4. Some others are discussed in Goodin 1995: ch. 8.

¹⁶ Ziegler and Tunney 2012 found that experimental subjects discounted future rewards more sharply when the beneficiary was more closely related to the decision-maker. That is, benefits to strangers were subject to less temporal discounting than those for cousins, which were discounted less than those for siblings, which were discounted less than those for the subject herself.

¹⁷ For a non-technical introduction to the formal results and the plausibility of the requirements see McLean 1987. On the relevance of social choice results to real politics see Mackie 2003.

¹⁸ On democracy and non-citizens see Abizadeh 2008. On democracy and future people see Mulgan 2006: ch. 8.

¹⁹ Asimov 1989 explores a limiting case of this possibility, in which elections are settled by interviewing a single person. I am grateful to Uri Leibowitz for drawing my attention to this wonderful short story.

can be justified on utilitarian grounds if the sole purpose of democracy is to elicit and aggregate information about what would promote utility.

But suppose we leave these problems aside and accept that democracy has some role to play in generating information about what would promote well-being, by providing policymakers and public officials with information about what people prefer. To the extent that this is true, it is no doubt important. However, note that it gives democracy a very narrow role in political life. On this picture, its role is limited to generating information. You say what you prefer, and I say what I prefer, and the election official notes down these preferences, which feed into a decision made elsewhere. While that is part of what happens in democratic politics, it is not all of what happens. In addition, you present me with reasons in favour of your preferred outcome or policy, and I try to counter them. Furthermore, you may seek to build a coalition of support for your position, and I may try to do the same for mine.

If we wish to do justice to these other aspects of political life we will have to depart from the purely technocratic models of politics that we identified earlier. One way to do this, I suggest, is to find some way of explaining the normative significance of considerations of legitimacy. This will enable us to explain one way in which political decisions do not simply involve the application of moral principles by experts. On the account I will give, the concept of legitimacy describes a certain kind of feasibility constraint—in particular, it describes which decisions or policies people would comply with for motives other than fear of sanctions. This conception of legitimacy is non-normative, but it has normative significance nonetheless. Among other things, it helps utilitarians to explain why democracy matters in ways that go beyond eliciting people's preferences.

8.3 Legitimacy and its Political Importance

We can begin our discussion of legitimacy by noting some features of our pre-theoretical concept of it. The aim is neither to give a complete catalogue nor to construct a theory that captures as many features as possible. That is not likely to be fruitful for utilitarians, since any account of legitimacy they offer is likely to be somewhat revisionary. Instead, the purpose is to avoid pure stipulation. Any theory that captures a reasonable number of the features of our pre-theoretical concept can claim to have some connection with that concept.

Our pre-theoretical concept of legitimacy is of a property of decisions, procedures, or institutions that is especially salient when there is disagreement about what should be done. Faced with such disagreement, we may say that if one party imposes her will on the others by force, that is illegitimate; whereas if she persuades them to agree with her, or they agree to settle the matter by voting, we may say that this is legitimate. If the issue arises in a context in which there is already a settled method for handling disagreements, we are likely to say that handling it using the established method is legitimate, but that advocating that some other method should be used to handle it is likely to be illegitimate. Notably, that seems likely even if the established method is no fairer, more just, or otherwise independently favoured than the alternative method. So far as legitimacy goes, we seem to give some priority to established methods over others, just in virtue of their being established.²⁰

It is not clear whether we take the primary bearer of the property of legitimacy to be the decision, the procedure, or the institution. Perhaps we apply the concept directly to more than one of these things. However, I shall assume that the primary application of the concept is to procedures, for simplicity. We can then hope to account for judgements of the legitimacy of decisions and institutions derivatively, taking procedures to be intermediate between decisions and institutions.

The salience of legitimacy when we face disagreements about what should be done helps to explain the importance of legitimacy in political philosophy. Such disagreements are surely central to political life, as Rawls and others have emphasized.²¹ We can think of legitimate procedures as mechanisms for coping with disagreements. They are not, primarily, mechanisms for *resolving* disagreements; they are an alternative to resolving disagreements. Since disagreements about what should be done are not confined to politics, however, our interest in legitimate procedures is not confined to politics, either. Children seem to latch on to the concept when making decisions amongst themselves, for example.²²

²⁰ Rawls 1999a: 318 comments on the priority of actual procedures. In a different context, so does Nozick 1974: 151–2.

²¹ Rawls 1996: 58–66 emphasizes that they may arise and persist among reasonable people. See also Miller 2016: 159 and Adams 2018: 87.

²² Political realists are committed to distinguishing quite sharply between politics and other parts of life. Though we hope to learn something from their emphasis on the distinctiveness of politics, we are not committed to any sharp distinction. Our explanation of the distinctiveness of politics will be that it usually exhibits to a higher degree a feature also found in other parts of life: that disagreement is, for practical purposes, ineliminable.

As these remarks indicate, we take legitimacy to be a normatively significant property. Whether or not some procedure is legitimate matters for how we should respond to the disagreement in question. Moreover, it seems to matter in a particular way. When a decision is arrived at legitimately, this seems to generate a reason to accept or comply with the decision that is at least somewhat independent of the content or nature of the decision made. This is most obvious in the case of fair lotteries: the reason to give the spoils to the winner is wholly independent of any other specification of who should get the spoils, since there is no other specification of that.²³ But it also seems true in cases where we do have independent standards for evaluating the decision. We may judge that the result of an election is legitimate but regrettable, for example, or that a superior's command is legitimate but mistaken.

We can use the concept of 'content-independent reasons' to describe this feature. These are reasons for action in which, as Joseph Raz puts it, 'there is no direct connection between the reason and the action for which it is a reason'.²⁴ The fact that an action would benefit someone is a content-dependent reason, since there is a direct connection between the reason and the action for which it is a reason. In contrast, the fact that a person with authority over me instructs me to do *X* is a content-independent reason for me to do *X*, since there is no direct connection between the reason—which is this person's authority—and the action for which it is a reason. If she had instructed me to do *Y* instead, the same reason—her authority—would have been a reason for that different action. Promises, commands, and requests all seem capable of generating content-independent reasons.²⁵

The legitimacy of procedures also seems to generate content-independent reasons. In the first place, when a legitimate procedure issues in a decision to do *X*, this generates a content-independent reason for the relevant agent or authority to do *X*. In the case of the lottery, it generates a reason to give the spoils to the winner, though there is no direct connection between the reason (the correct implementation of the lottery procedure) and this person's getting the winnings. Second, it may generate content-independent reasons for others to comply with or go along with the decision. The legitimacy of an

²³ Here I assume that the lottery is a matter of 'pure procedural justice', in Rawls's sense. He writes: 'pure procedural justice obtains when there is no independent criterion for the right result: instead there is a correct or fair procedure such that the outcome is likewise correct or fair, whatever it is, provided that the procedure has been properly followed' 1999a: 75.

²⁴ Raz 1986: 35. The concept of content-independent reasons is due to Hart. See Hart 1958: 100–2 and Hart 1982: 254–5.

²⁵ Raz 1986: 35–7. Raz notes that threats and offers also generate content-independent reasons—though he claims that these are reasons for belief rather than for action.

election generates a reason for members of the electorate to accept its result that is independent of who wins the election and their merits.

The features of our pre-theoretical concept of legitimacy that we have identified are that it is a property of procedures that is salient in cases of disagreement, and that it has normative significance in virtue of generating content-independent reasons. Moreover, as we have noted, legitimacy does not imply rightness. It is consistent, and common, to claim that a legitimate procedure has issued in the wrong decision. This is, indeed, the standard combination of attitudes amongst democrats who find themselves on the losing side of a vote. In such cases we may believe that the legitimacy of the procedure generates a reason to accept the decision, even though we judge the decision to be wrong.

In this way, judgements of legitimacy are relatively independent of judgements of rightness. In fact, this independence is essential for legitimacy to play its characteristic role of being a backstop method for dealing with disagreement. If we could come to agree about what is right, we would have little need for the concept of legitimacy. Since we sometimes cannot, for practical purposes, come to agree about what is right, we rely instead on the concept of legitimacy. Legitimacy could not play this role if it were tightly conceptually connected to rightness.

As a matter of political psychology, then, the fact that we can judge a decision to be legitimate but morally wrong is vitally important. This is borne out by the psychologist Tom Tyler's research into why people comply with laws.²⁶ Tyler distinguished three broad kinds of motivation: incentives (including the threat of punishment), concordance with personal moral views, and a sense of legitimacy. Incentives are costly and difficult to implement, while in a diverse society we cannot expect all laws to be in accordance with the personal morality of all or most people. The sense of legitimacy is therefore important in motivating compliance with the law.

However, the sense of legitimacy also has its dangers. Because it can motivate people to comply with the law even when they could gain something personally by breaking it, or they do not agree with it on moral grounds, it grants discretion to leaders. Tyler writes:

Leaders are especially interested in having legitimacy in the eyes of their followers, because legitimacy most effectively provides them with discretionary authority that they can use in governing. Morality can lead to

²⁶ See Tyler 1990. See also Tyler and Mitchell 1994, which examines the role of judgements of legitimacy in the context of the deeply morally divisive issue of abortion, and Tyler and Darley 2000.

compliance with laws, but it can also work against it. For example, during the war in Vietnam those who believed in the legitimacy of the government fought in the war regardless of their personal feelings about its wisdom. For others the perceived immorality of the war was a factor leading them to oppose and violate the law.²⁷

As Tyler notes, the relative independence of judgements of legitimacy from judgements of rightness makes legitimacy useful in accommodating disagreement. It provides a way of generating agreement about how to proceed that does not depend on resolving underlying moral disagreements. However, it also makes the sense of legitimacy open to abuse.

It is an important fact about humans that we can be motivated to comply with a decision, policy, or law because we judge it to be legitimate, even when we judge it to be wrong. It means, on one hand, that we can appeal to considerations of legitimacy as a way of coping with disagreement. Where we must cooperate but we cannot agree on what is right, we might be able to proceed based on agreement about legitimacy. On the other hand, the fact that we can be motivated to comply with a decision that we judge to be wrong can be exploited or abused by leaders.

8.4 Utilitarianism and Legitimacy

It may seem difficult for utilitarians to account for the significance of legitimacy. One apparent difficulty is the mismatch between utilitarianism's focus on outcomes and the fact that legitimacy appears to be a property of procedures. Whether a decision is legitimate depends on its provenance, or how it was taken, while whether it is right depends on what it causes, according to utilitarianism.²⁸

This is not such a difficult problem as it may first appear. Since utilitarians can evaluate procedures, they can in principle try to specify properties of procedures that make them legitimate. If they can do that, they can characterize the legitimacy of decisions in terms of their provenance. The

²⁷ Tyler 1990: 4.

²⁸ Pettit 1997 claims that what matters for the avoidance of 'arbitrariness' in political decisions is that decisions are contestable, not that they arose in a specific way. Contestability is a modal, not an historical, property (185). However, contestability requires the existence of certain procedures (187–90). Hence, even on Pettit's view there is an apparent mismatch between the focus of questions of legitimacy (insofar as that requires 'non-arbitrariness') and utilitarianism's focus on outcomes.

legitimacy of decisions would then be an indirect matter, having to do with whichever properties of procedures are said to determine their legitimacy. Although it remains to be seen whether there is a plausible way for utilitarians to characterize the legitimacy of procedures, the mere fact that the legitimacy of decisions depends on their provenance is not sufficient to show that utilitarians cannot account for it.

However, there is a harder problem. We have said that legitimacy is salient in cases of disagreement, and that it appears to be a mechanism for coping with it. The harder problem for utilitarians is to find an appropriate role for legitimacy within their theory. This problem is made harder because of the way most political philosophers think of legitimacy. Most treat it as a normative property—that is, as a property which cannot be specified without using normative concepts such as ‘reason’ or ‘rightness’. So-called ‘moralists’ claim that it is a moral property, though they disagree about which moral property.²⁹ So-called ‘political realists’ often deny that it is a moral property, claiming instead that it is a specifically political property—but they agree nonetheless that legitimacy is a normative property.³⁰

Thinking of legitimacy as a normative property of some sort makes it harder to see how utilitarians could account for its significance, since it seems to imply that legitimacy must be either a contributor to rightness or a distinct kind of normative status on a par with rightness. The idea that legitimacy is a distinct kind of normative status would make sense if we were to divide life into different realms of activity, with rightness governing one realm (say, the moral) and legitimacy governing the other (say, the political). But utilitarians are not likely to accept that sort of picture. Without some such division, we would not know whether to do the legitimate thing or the right thing in cases where they conflict. But it is equally hard to see how utilitarians could regard legitimacy as a contributor to rightness: as a right-making feature of decisions. For then its normative significance would presumably be absorbed in its contribution to rightness, and it is hard to

²⁹ Allen Buchanan defines ‘political legitimacy’ as possession of moral justification in the ‘making, application, and enforcement of laws’ 2003: 146. Estlund 2008: 41 defines legitimacy similarly as permissibility in the state’s use of coercive power. These authors distinguish sharply between legitimacy (permissibility in issuing and enforcing commands) and authority (which implies a duty to obey): see also Edmundson 1998. In contrast, Raz 1986: ch. 2 uses ‘legitimate’ to refer to a particular kind of authority, which implies a duty to obey.

³⁰ For example, see Rossi 2012. Williams 2005: 10 rejects the idea that political legitimacy is a matter of ‘mere moral normativity’. This leaves open whether it reflects a special kind of moral normativity, or some non-moral kind of normativity. On the idea of political normativity see Maynard and Worsnip 2018.

see how it could play the role of enabling cooperation when we disagree about rightness.³¹

However, there is nothing to stop utilitarians finding a role for legitimacy if they think about it in a different way. In the rest of this chapter we will investigate the normative significance of legitimacy understood as a non-normative property. A procedure or decision can be legitimate, in the sense to be defined, without meeting any normative standard. Like many other non-normative properties, however, legitimacy in this sense has considerable normative significance. In fact, I shall argue that it has enough normative significance to greatly enrich the utilitarian account of politics.

My proposal is that we should define the legitimacy of a procedure as the degree to which those subject to it comply with its decisions for reasons other than threat of sanctions for non-compliance. On this account, legitimacy is a sociological fact about a procedure.³² It is the degree to which, in fact, people subject to the procedure comply with it out of motivations other than avoidance of sanctions. Thus only actual procedures, strictly speaking, can possess legitimacy—though we can also discuss the degree to which some hypothetical procedure would be legitimate, if it were realized.

Legitimacy in this sense is more than bare Hobbesian security. Bare Hobbesian security certainly has significant advantages when compared with insecurity. But not all regimes that produce Hobbesian security are also legitimate in the sense I have described, since some of them rely to a high degree on coercion in producing security. On the other hand, legitimacy in the sense I have described does not require satisfaction of any normative standard, such as consent, justice, or fairness. A procedure could be legitimate yet unfair and unjust, and it could be legitimate despite the fact that those subject to its decisions have not consented to it. All that is required is that they comply for reasons other than fear of sanctions. Their compliant behaviour could, for example, be due to habit, lack of attention, or false consciousness, just as it could be due to belief in the procedure's

³¹ This may explain the apparent neglect of the topic by utilitarians. Binmore 1998 is one of very few utilitarian accounts of legitimacy. Moreover, he appears to understand legitimacy as closely connected with fairness, and he claims that a concern with fairness, properly understood, leads to utilitarian conclusions. For him, concern with legitimacy is thus an alternative way to reach utilitarian conclusions about rightness. Sidgwick 1907: bk. 3 ch. 6, Simmons 1979: 45–54, Greenawalt 1987: ch. 6, and Hare 1989: ch. 2 discuss utilitarianism and the related topic of political obligation.

³² My proposal is broadly in line with the account given in Weber 1994b. Note that my definition is broader than Tyler's, since it does not distinguish compliance due to a sense of legitimacy from compliance due to concordance with moral beliefs.

fairness or justice. Legitimacy in this thin sense is less demanding not only than moralists' concepts of legitimacy, but also than Williams's realist concept. Nor need it involve obedience as that is usually understood: it is merely compliance for reasons other than coercion.³³

Though it is not a normative property, legitimacy in this thin sense has considerable normative significance, and its interest depends on normative matters. Many non-normative properties have normative significance: weight, for example, is normatively significant when you are considering whether to cross a rickety bridge. The degree to which people comply with decisions for reasons other than fear of sanctions is also normatively significant. Our interest in this non-normative property reflects the significance of the costs of coercion and disorder.

For a utilitarian, the legitimacy of a procedure in the sense I have just described is likely to matter greatly. Legitimacy is not guaranteed to be present, and may sometimes be impossible: in many circumstances, order is possible only through coercion. Where legitimacy exists, it is a background feature of circumstances; if we know that some procedure or institution is legitimate, we know something about what people on the whole would or would not accept. Legitimacy is in this way a constraint on feasible non-coercive action. Its normative significance flows from its implications for what can be achieved non-coercively. Because of the costs of coercion, both to the coercer and to the coerced, and the costs of insecurity and disorder, utilitarians will tend to think that legitimacy is usually of great value.

Where legitimacy in the sociological sense does exist, it is the result of convergent *de facto* acceptance of some procedure. This provides one possible explanation of the observation we made earlier, that in matters of legitimacy priority tends to be given to established procedures over merely possible ones. Where there is a legitimate procedure, it would typically be regarded as illegitimate for someone to propose shifting to some other procedure for the purpose of settling a matter under current dispute, even if that alternative procedure would be fairer or superior in some other way. To proceed legitimately we need first to secure acceptance of the new procedure in a way that is independent of its implications for any current

³³ Here I understand coercion to be the modification of behaviour through use of credible threats. For discussion of different conceptions of coercion, see Wertheimer 1987 and Anderson 2008.

matter of dispute. This conservative aspect of legitimacy is just what we would expect on the account I have proposed. To achieve convergent acceptance of a procedure it is likely to be necessary to restrict the range of salient alternatives, which is what the tacit requirement to secure acceptance independently of any current matter of dispute achieves.

Moreover, legitimacy in this thin sense will give rise to content-independent reasons. The fact that some decision arose from a legitimate procedure will, normally, imply other people's acceptance of it, which will in turn affect the consequences of an agent's compliance or non-compliance with it, and so affect her reasons for or against compliance. Generally, it will make the consequences of non-compliance worse on utilitarian grounds than they would have been had the decision not been legitimate and so not widely accepted. Furthermore, to the extent that widespread acceptance of a procedure is beneficial, there may be an additional pattern-based reason to contribute to the production of these benefits. In these ways, legitimacy usually produces one or more additional reasons for compliance. These reasons flow from the general acceptance of the procedure, rather than from the independent character of the decision produced by the procedure. We can see this by noting that if the procedure had reached a different decision, the very same reasons for compliance would have applied to that.

The fact that a decision is legitimate in the thin sense I have described is, of course, perfectly consistent with the decision's being wrong. Further, in common with other accounts of legitimacy, our account implies that the legitimacy of a procedure or political regime does not entail a duty to obey.³⁴ It does imply that in most cases there will be a content-independent reason to comply with the decision, but this falls short of a duty to obey as that is usually understood. The ordinary idea of a duty to obey includes several elements, and one of these is the idea that the reason to obey is sufficient to make non-compliance almost always wrong.³⁵ Our account of legitimacy implies that there is usually a content-independent reason to comply, but it does not imply that failure to comply is almost always wrong. Recall that a

³⁴ For example, see Edmundson 1998, Wellman 2001: 741, Buchanan 2003: 146, and Estlund 2008: 41.

³⁵ Raz 1986: 60. Raz's explanation of this idea depends on what he calls the 'pre-emption thesis', which is the claim that the reason to obey an authoritative directive replaces rather than supplements the reasons the agent has independently for or against complying with the directive (59). The account of legitimacy offered here does not imply Raz's pre-emption thesis. More generally, it does not aspire to being an account of authority or of political obligation.

procedure may be legitimate, in our thin sense, yet deeply unfair or unjust. In such cases—and even where the procedure is not only legitimate but also just—it may be right to do something other than comply. For example, it might be right to protest, or to campaign for reform, or to be civilly disobedient. The fact that a procedure is legitimate will almost always make a difference to an agent's reasons, but it will not almost always make it wrong to fail to comply with the decisions it generates.

We might ask whether this non-normative property of legitimacy has enough normative significance, or significance of the right kind. It does seem to explain those features of our pre-theoretical concept of legitimacy that we identified in Section 8.3. It specifies something with normative significance; it is especially salient in cases of disagreement about what should be done; and it generates content-independent reasons to comply with legitimate decisions even when we judge them to be wrong. This answers the challenge of showing how utilitarians can account for the significance of legitimacy. On the other hand, we did not attempt to specify all of the features of our pre-theoretical concept of legitimacy. Most obviously, there is no normative standard that a procedure must meet in order to be legitimate on our account. This means that a procedure could in principle be legitimate in this sense while being as bad as we can imagine in other respects. This makes the thin account undeniably revisionary of our pre-theoretical concept.

Utilitarians could specify further requirements for legitimacy. They could claim, for example, that legitimacy is a complex property of procedures, consisting of (a) being accepted in the way we have specified, and (b) passing some normative threshold, such as a threshold of justice or democracy. In that way, utilitarians could hope to recapture some of the normative content of our pre-theoretical concept of legitimacy that is not included in the thin account proposed here.

On the other hand, we already have ways to refer to the normative properties that might be specified under (b). We do not need to build a threshold of justice or democracy into the concept of legitimacy in order to discuss or evaluate the justice or democracy of a procedure or institution. We can simply distinguish clearly between legitimacy, justice, and democracy. On balance, it seems to me that the gains in simplicity and clarity that we get from sticking with the thin account outweigh the loss involved in its being revisionary. The thin account identifies a specific feature that some procedures have and others lack, and which for utilitarians has considerable normative significance. There is something to be said for isolating this

feature rather than bundling it together with other desirable features of procedures.³⁶

8.5 Should Utilitarians Be Democrats?

Now let us return to consider the utilitarian case for democracy, and see whether considerations of legitimacy add anything to the picture. Earlier we noted the limitations of the utilitarian argument in favour of democracy that is based on the idea that democratic procedures elicit and aggregate voters' preferences. Of course, utilitarians have offered other reasons in favour of democracy. Let us briefly review these, and ask whether considerations of legitimacy make a significant difference to the utilitarian case for democracy.

It is common and helpful to distinguish between the aggregative conception of democracy, which we have already considered, and a so-called 'deliberative conception'.³⁷ On the deliberative conception democracy takes place between elections as well as at them, and citizens are expected to participate in democratic processes by offering reasons or expressing views about the common good, rather than by simply stating their preferences.³⁸ By expecting each citizen to do more than provide information about what would be good for her, the deliberative conception gives less emphasis to the idea that each is the best judge of her own interests. It may nevertheless hope to gain insights from the different perspectives and social positions that people occupy.³⁹

Utilitarians can endorse the deliberative conception of democracy, at least insofar as the deliberation is about what would be good for people or the means of promoting well-being. Public debate and voting can bring to light both new information about what people would prefer and new insights about what would be for the common good. As Rawls says, '[d]iscussion is a way of combining information and enlarging the range of arguments. At least in the course of time, the effects of common deliberation seem bound to

³⁶ Adams urges 'minimalism' about legitimacy, since '[t]he more we build into legitimacy, the more controversial it becomes, and so the less able it is to play its distinctive role' 2018: 88.

³⁷ See, for example, Miller 1992.

³⁸ See, for example Cohen 2002. Like the aggregative conception, the deliberative conception of democracy faces theoretical problems in spelling out its core idea. Chief among these is the so-called 'discursive dilemma'. For discussion of this problem see Pettit 2001.

³⁹ This is emphasized by Iris Marion Young. See Young 2002 and Peter 2007.

improve matters'.⁴⁰ Democratic discussion is also a way of achieving diversity amongst decision-makers, which for quite general reasons has its advantages.⁴¹ One of its most important functions may be to guard against a danger of rule by experts: exaggerating the importance of the sorts of considerations that are trackable by experts.⁴²

A different kind of utilitarian argument in favour of democracy has to do with the dispersal and constraint of power. This theme is especially prominent in Benthamism, perhaps because of its acceptance of psychological egoism. If every individual is always motivated by what he takes to be his own interest, the problem of finding ways to disperse or constrain power, so as to avoid the obstruction of the common good by 'sinister interests', is particularly pressing.⁴³ However, we do not have to be psychological egoists to expect those who hold power sometimes to abuse it in pursuit of their own ends. It is an empirical matter what the utility-maximizing distribution of power in a society would be, but common experience suggests that, to the extent that it involves concentrating power in some hands, it will also involve safeguards against the misuse of that power. Some utilitarian reasons for supporting democracy come from this sort of consideration.⁴⁴

One kind of safeguard is to try to align the interests of those holding power with the common good. This is the guiding idea behind James Mill's advocacy of frequent elections. If the governors can expect pretty soon to be among the governed, they may well make laws that suit the governed and not just the governors.⁴⁵ Note that there are some reasons to think that this safeguard would not work properly if the electorate were merely a

⁴⁰ Rawls 1999a: 315.

⁴¹ Scott Page has analysed the benefits of diversity in complex systems generally (2011). He notes that diversity provides a sort of insurance against shocks, as in diversified investment portfolios (ch. 6). He also notes that there may be diminishing marginal utility from each member of any given type, so that greater diversity of types tends to promote utility (ch. 7). He summarizes: 'If we have enough information to know what drives performance, then we should select the best collection on the basis of that information. If not, and if we only get one try, then we should probably choose a diverse collection. This does not mean that diversity is always better, only that if we are not sure of what we're doing, we should err toward greater diversity' (114). It is quite plausible that, when we face political choices, we often are not sure of what we are doing.

⁴² Williams writes: 'Just as in the natural sciences, scientific questions get asked in those areas where experimental techniques exist for answering them, so in the very different matter of political and social decision weight will be put on those considerations which respected intellectual techniques can seem, or at least promise, to handle' 1973: 148. Williams presents this as a criticism of utilitarianism, but it might be seen instead as a criticism of certain technocratic applications of utilitarianism.

⁴³ See Schofield 2006: ch. 5 and Mill 1978/1820.

⁴⁴ On safeguarding mechanisms in general see Elster 2013.

⁴⁵ Mill 1978/1820.

representative sample of the population—for then they could become a sectional interest in their own right. As a result, the logic of this argument for frequent elections provides some reason for extending the franchise to all of those whose interests are affected.⁴⁶

A second way in which democratic processes can be a safeguard against misuse of power is that they can act as a remedial device when things do not work as hoped. Elections are, among other things, a mechanism for removing from power those who use it poorly. This is valuable independently of the incentive it gives to rule well while in office, since it is a safeguard against incompetence, not just against sinister interests. It is valuable also as a safeguard against the corrupting effects of being in power. Even if the new ruler starts out no better than the old, in motivation or competence, there is something to be said for preventing any individual or group from maintaining a grip on power for a long period. As the Leveller John Lilburne put it, ‘Suffer not one sort of men too long to remain adjutators, least they be corrupted by bribes of offices, or places of preferment, for standing water though never so pure at first, in time putrifies’.⁴⁷

In addition to eliciting and aggregating preferences, then, democratic processes may promote utility by enabling public deliberation and by acting as safeguards against the misuse of power. These arguments have in common the idea that democracy promotes utility by generating better laws, policies, or decisions. A quite different utilitarian rationale for democracy was offered by John Stuart Mill. He wrote:

It is not sufficiently considered how little there is in most men’s ordinary life to give any largeness either to their conceptions or to their sentiments. Their work is a routine; not a labour of love, but of self-interest in the most elementary form, the satisfaction of daily wants; neither the thing done, nor the process of doing it, introduces the mind to thoughts or feelings extending beyond individuals; if instructive books are within their reach, there is no stimulus to read them; and in most cases the individual has no access to any person of cultivation much superior to his own. Giving him something to do for the public, supplies, in a measure, all these deficiencies.

⁴⁶ This did not prevent James Mill from recommending that women be excluded, on the ground that for ‘almost all’ of them their interest ‘is involved either in that of their fathers or in that of their husbands’ 1978/1820: 79. If nothing else, this illustrates the importance of factual assumptions for utilitarians.

⁴⁷ Cited in Elster 2013: 6.

If circumstances allow the amount of public duty assigned him to be considerable, it makes him an educated man.⁴⁸

Mill's idea is that democracy has a beneficial effect on those who participate in its processes, quite apart from any effect it may have on the quality of political decisions.

Of course, Mill is here speculating on an empirical issue. It is plausible that involvement in political matters—in voting, discussion, or service in local or national public office—would have some effect on people's character and knowledge. It is not completely clear, however, whether the effect would usually be an improvement, still less whether it would be a sufficient benefit to offset the costs of involvement. We can speculate about these matters, but we need empirical evidence to settle them. There is some evidence from political science that increased involvement in political processes—in particular, involvement in referenda and other forms of direct democracy—increases voter knowledge of political information.⁴⁹ However, this evidence is only obliquely relevant to Mill's hypothesis, since it concerns a narrow kind of political knowledge and not education in Mill's broader sense.⁵⁰

As this brief review suggests, the utilitarian case for democracy is a patchwork of overlapping instrumental considerations. Since they are instrumental, they depend on empirical claims.⁵¹ Unfortunately we do not always have sufficient evidence for these claims, and so the instrumental case has to be somewhat speculative and tentative. On the other hand, a *prima facie* case for them can be made by examining the record of democracies as compared with other forms of government.⁵² The so-called 'democratic peace thesis', according to which democracies do not wage war with each other, is one piece of evidence of the benefits of democracy.⁵³ Amartya Sen's thesis that democracy protects against famine is another.⁵⁴

⁴⁸ Mill 1991c/1861: 254. Arneson 2003: 123 also notes the importance of this possible effect of democracy.

⁴⁹ See Benz and Stutzer 2004 and Smith and Tolbert 2007.

⁵⁰ Pedersen 1982 surveyed evidence for and against Mill's hypothesis. He found little convincing evidence either way, concluding that 'the causal connection between political participation and growth of conscience . . . must remain an article of faith' (566).

⁵¹ Sidgwick describes the conflict between epistocracy and democracy as 'a conflict of which it is impossible to find a solution, so long as the argument remains in the *a priori* region' 1907: 299, italics in the original.

⁵² Runciman 2013 is an interesting study of the tendency of democracies to find themselves in crisis, and of the ways in which they respond to them.

⁵³ Doyle 1983 is an influential statement of the democratic peace thesis. Rosato 2003 criticizes it.

⁵⁴ See Sen 1999: 16, 51–3, 155–7.

As well as being tentative, the utilitarian case for democracy is limited. Utilitarians will want to take the costs and likely disadvantages of democracy into account alongside its likely benefits. As we noted, one very significant problem with democracy for utilitarians is that it does not appear well-suited to looking after the interests of non-voters, including foreigners and future people. Another problem arises in cases where issues are highly technical and require expert judgement. Where these problems intersect, as they do in regard to climate change, we might expect democracies to take poor decisions.⁵⁵

As a result, utilitarians are likely to seek measures to ameliorate democracy's failings. Such measures might include constitutional constraints on democratic decisions, and finding an appropriate role for expertise in the political process. They are also likely to include wanting to cultivate the right sort of political culture and the right sort of civic virtue. In order to promote utility, democracy has to be embedded in a suitable wider environment. A further implication of utilitarianism's instrumental commitment to democracy is also, of course, that it would be withdrawn in circumstances where democracy would not promote well-being.

However, one final consideration suggests that these circumstances may now be very rare. Arguably, democratic processes are now a prerequisite of legitimacy in much of the world. This is not because, as a matter of principle, legitimacy requires democracy. On the account of legitimacy I have proposed, the connection between legitimacy and democracy is not a logical one. Instead, legitimacy may require democracy for the practically very significant reason that many people now would not accept any alternative to democratic decision-making, at least for some kinds of decision. If political decisions in their society were taken undemocratically, many people would not comply with them except out of fear of sanctions.

This is not the sort of rationale for democracy that will seem important to those who think of democracy as a requirement of justice or as an expression of the fundamental equality or dignity of citizens. It is not inspiring. Nevertheless, it surely adds significantly to the reasons that utilitarians have for supporting democracy.

⁵⁵ See Mulgan 2011: pt. 4 and Jamieson 2014: 100–1.

8.6 Conclusion

The thin account of legitimacy enables us to explain one way in which politics is somewhat different from other parts of life. In private life we can often try to resolve disagreement through discussion, or to avoid it by associating only with others who are like-minded. The special character of politics is due to the fact that these methods are not practicable on a large scale. No one can hope to persuade all of her fellow citizens to agree with her about moral matters, nor to form a political community only with those who are like-minded. We can try to engineer bubbles of agreement to inhabit in our private lives, but we cannot hope to do that in politics. In politics, disagreement is more stubborn and more pressing than, at least some of the time, it is in private life.

This distinction is not sharp, but is instead a matter of degree. Yet it explains why legitimate procedures have a special importance in politics. We resort to considerations of legitimacy when we cannot resolve disagreement, but instead have to find some way to live with it. On the account I have proposed we can think of legitimacy as a special kind of feasibility constraint, since it describes the ways in which order can be achieved without relying on coercion. Unless the circumstances are such that the costs of coercion or disorder are worth paying, the pursuit of utilitarian aims by political means is thus constrained by considerations of legitimacy. Where democracy is a prerequisite of legitimacy, it too constrains the pursuit of utilitarian aims.

This provides an attractive way of understanding the distinctiveness of politics. On this analysis, politics is not distinctive in the way that Rawlsians or political realists claim. It is not subject to different kinds of reason or distinct kinds of normative principle. Instead, it is distinct to the extent that the constraints on the pursuit of our goals are different, and more pressing, in politics than in other parts of life. The conditions of widespread disagreement characteristic of politics impose a special kind of feasibility constraint on political decisions. It is this feasibility constraint that should lead us to reject the purely technocratic models of politics, rather than any deep discontinuity of underlying principles.

Virtuous Agents

The concepts of reasons and rightness that we have been discussing are non-perspectival. This means that an agent's reasons, and what it is right for her to do, depend on the way the world really is: they do not depend on how it seems to be from any perspective, including hers. If my well-intentioned action would just make things worse, then I have no reason to do it, in this sense of 'reason', no matter how justified my beliefs.

Since we are considering utilitarian theories, perspectives do still play a role in this picture. A person's well-being presumably depends in some way on how things seem to her. If I do not understand that his remark was an insult directed at me, some of its impact on my well-being must be undermined. So perspectives help to fix the facts about well-being. But once those facts are fixed, perspectives play no other role in determining what reasons anyone has, or what it would be right for anyone to do, so far as the non-perspectival concepts of reasons and rightness are concerned. In particular, reasons and rightness depend on the truth about the consequences of actions, about which actions are possible, and about their relationship to patterns of action—rather than on beliefs about these things.

This makes reasons and rightness highly detached from our day to day life and thought. Because utilitarians count equal-sized gains or losses in well-being as equally important no matter when they occur, it is in principle the whole future that matters. If reasons and rightness depend on how that future turns out rather than on any beliefs about it—no matter how reasonable or well-justified they are—it is difficult to see how reasons and rightness, in this non-perspectival sense, can play any role in the ethical life of agents.

This sets up a double challenge for the version of utilitarianism I have been outlining. The first challenge is to explain what, if anything, this version of utilitarianism can say about several of the things that loom large in our perspectives on ethical matters. These include the characteristics of good decisions or good decision-makers, what it takes to be praiseworthy or blameworthy, and what it takes to be virtuous or vicious. It is hard to

accept that the answers to these questions are also non-perspectival. We want an account of good decision-making to explain how an agent, inhabiting a certain perspective on the world, can reach a good decision. We do not want an account of praiseworthiness or virtue to find fault with an agent who has only justified beliefs, just because some of those beliefs turn out to be false. Our judgements on these matters seem attuned to how things seem from agents' perspectives. But if that is correct, the accounts of good decision-making, praiseworthiness, and virtue cannot be connected in any very simple way with the accounts of non-perspectival reasons and rightness.

The second challenge is to show that, if we can give some account of good decision-making, praiseworthiness, and virtue, the non-perspectival concepts of reasons and rightness will not be drained of their significance. If good decisions are not defined as ones that are based on justified beliefs about rightness, for example, why should we care about rightness? If virtue is not a matter of responding appropriately to reasons, why care about reasons?

The aim of this chapter is to answer these challenges. Now, a natural thought is that giving a plausible account of good decision-making, praiseworthiness, and virtue will require introducing perspectival concepts of reasons and rightness—so that we can say, for example, that acting in a way that is wrong in a perspectival sense is blameworthy, or that the upshot of good deliberation is a judgement about which action is right in a perspectival sense. That would be to accept perspectival/non-perspectival pluralism, supposing that we accept the argument of Chapter 3 that we should retain the non-perspectival concepts of reasons and rightness.

However, in this chapter I will resist this pressure towards pluralism, and claim instead that it is possible to give plausible utilitarian accounts of good decision-making, praiseworthiness, virtue, and their opposites without introducing perspectival concepts of reasons or rightness. Moreover, I shall claim that the *non*-perspectival concepts of reasons and rightness have a vital role in the accounts of these matters. Even though good decision-makers do not deliberate about non-perspectival reasons or rightness, on my view, what makes them good decision-makers is that they act in ways that are right in the non-perspectival sense more often than do others. In other words, the non-perspectival concepts provide the criterion of good decision-making, which in turn plays a role in the accounts of praiseworthiness and virtue. Thus, the non-perspectival concepts of reasons and rightness are highly significant even though they are not directly action-guiding.

These issues have particular importance for utilitarians. Along with other consequentialist theories, utilitarianism has frequently been accused of leaving insufficient room for partial concerns and attachments.¹ Since it is in many respects a highly impartial theory, this kind of criticism has a good deal of initial plausibility. One important line of response is to try to show that the virtuous utilitarian agent will have partial concerns and attachments; that, despite his reputation, he is capable of being a devoted brother or a loyal friend.² To give that response, utilitarians need first to show that they can give a plausible account of the characteristics of virtuous agents.

9.1 Reasons and Rightness

Neither virtuous agents nor good decision-making aim directly at acting rightly in the non-perspectival sense. As I shall explain, we have so little idea about which token actions are right in this sense that aiming to identify them is, quite generally, a poor way to live.

Nevertheless, it does not follow that virtue and good decision-making have nothing to do with the non-perspectival concepts of rightness and reasons. The accounts of virtue and good decision-making that I will offer in this chapter are instrumentalist, in the sense that they characterize both virtue and good decision-making in terms of causal relationships between psychological features of persons and right decisions, in the non-perspectival sense of 'right'. Because they take this instrumentalist form, it makes sense first to consider what the account developed in this book has said about non-perspectival reasons and rightness. In this section I briefly recapitulate its main features, and point out the ways in which it is still incomplete. Despite these gaps, though, we know enough about its broad structure to employ it in the accounts of good decision-making and virtue discussed in the rest of this chapter.

The pluralist form of utilitarianism we have been exploring recognizes two kinds of reasons. Act-based reasons are generated by the causal properties of actions. When an action would prevent some harm or produce some benefit, we have an act-based reason to do it. Pattern-based reasons are generated by the parthood properties of actions. When an action would be

¹ For example, see Williams 1973: 93–118, Wolf 1982, and Scheffler 1994: ch. 3.

² Smart writes: '[o]ne must not think of the utilitarian as the sort of person who you would not trust further than you can kick him' 1973: 71.

one's part in some valuable and eligible pattern of action, we have a reason to do it. Both kinds of reason are associated with specific benefits and harms, rather than overall goodness. Thus, for example, Jack's pleasure today may give rise to one reason, while Jane's pleasure today or Jack's pleasure tomorrow give rise to others; and these reasons may conflict with each other. All reasons are to be explained in terms of goodness, but it does not follow that reasons are to be explained in terms of what is best, all things considered.

I claimed that recognizing both kinds of reasons enables utilitarians to explain much of the complexity of common moral judgements. When we have misgivings about Act Consequentialism, it is because we think it misses some reasons that exist, not because we think the reasons it claims exist do not. The fact that an act would make someone better off is always a reason in its favour. Act Consequentialism is, as I put it, half right. Chief among the reasons we think it misses out are moral constraints, including the reasons that constitute moral rights. Moral constraints are agent-relative: they are reasons for each agent to see to it that he does something or that he refrains from doing something. To explain this feature using a theory of value, we have to assign extra value (or disvalue) to the agent's doing this thing now, as compared with his doing it at some other time or others' doing it. However, these claims about value are highly implausible when considered in themselves.

It is much more promising to try to explain moral constraints using the idea of pattern-based reasons. Pattern-based reasons are naturally agent-relative: they are reasons for each agent to play his part in a pattern of action. The parthood relation singles out the agent's action, while the causal relation at work in act-based reasons does not. My act-based reasons regarding torture are to act so as to reduce it; but this is compatible with torturing someone myself. If we want to explain a moral constraint against torture, we need a reason that attaches specifically to any possible act of torture I may commit. Assuming that a pattern of action in which no-one tortures is eligible and valuable, I have a pattern-based reason to play my part in it. This reason attaches specifically to my part, which is an action that only I can perform. It has exactly the right form to be a moral constraint.

However, pointing out that a concept has the right form to play a role in explaining something does not amount to explaining that thing. I have tried to be clear about the gaps in the explanation of moral constraints that I have offered. To reiterate, these include the following. First, I have not offered a theory of goodness. I have claimed that welfarism is a plausible hypothesis;

I have also made several claims about the circumstances under which someone could know what is good for her, but I have not offered a theory of the nature of well-being. Second, I have not offered a full account of the eligibility of patterns. In light of Dietz's criticism of my earlier views, I accepted a weak version of the willingness requirement, according to which a pattern of action is eligible only if there is some chance it would be performed were the agent to play her part in it. But as I explained in Chapter 5, we will need to restrict eligibility further in order to explain moral constraints, and I have not identified further constraints on eligibility. Third, claims about what pattern-based reasons there are depend on empirical claims about the existence and consequences of patterns of action. These claims are highly uncertain.

Suppose, for the sake of argument, that these gaps were successfully filled. We would then have a theory which tells us which reasons exist. It would not yet tell us which actions are right, however. To do that, we would need two more things. The first is a theory of the strength or weight of reasons. This would tell us the strength of each reason, and so would also tell us which reason is strongest in any case. To get from there to a theory of rightness we would need, finally, a theory of the function from reasons to rightness. This could say, simply, that the right action in each case is the option for which there is strongest reason.³ But things need not be as simple as that. For example, another candidate function is this: an action is right if and only if the agent has a moral right to perform it or it is the option for which there is strongest reason.⁴

These two functions yield different implications when the agent exercises a right by performing an action for which there are weaker reasons than some alternative—for example, by spending money to which she is morally entitled on a trip to the cinema rather than on famine relief. The simple function implies that this action is wrong, while the alternative function implies that it is right. Either view is, so far as I can tell, compatible with the commitments of utilitarianism. Other functions are, no doubt, also possible.⁵ Obviously, it matters a great deal which of these functions we adopt.

³ Or, in the case of ties, an option for which there are reasons at least as strong as those for any alternative. For ease of exposition, I ignore this complication here.

⁴ 'Or' here is inclusive.

⁵ Dancy 2004: chs. 2–3 emphasizes the complex ways in which reasons may interact.

When judging the relative strength of reasons, it might be thought that pluralism about kinds of reason poses especially intractable theoretical problems. We might suppose that we cannot weigh up the strength of act-based reasons against pattern-based reasons, because they are of such different kinds.⁶ We cannot, for example, calculate the strength of a pattern-based reason by asking what difference the action would make to the outcome—for then we would be evaluating it as if it were an act-based reason. However, I do not see any special difficulty here; these matters are difficult in the ordinary way that ethical judgements are difficult. We can theorize about plural reasons in the same sort of way that we might theorize about plural values. Typically this will involve careful consideration of different cases, and trying to spot patterns in our judgements about them. I assume that, here as elsewhere, an appropriate method is the search for reflective equilibrium.⁷

One constraint on this theorizing can help to guide us. Earlier we noted that act-based reasons are a special case of pattern-based reasons. This means that we are really searching for a single theory of the strength of reasons, with acceptable implications about both act-based and (other) pattern-based reasons. We can illustrate the potential significance of this constraint with an example. Suppose that we adopt the following simple theory of the strength of pattern-based reasons:

STR_{PBR} The strength, S , of agent A 's reason to play her part, X , in a pattern, P , performable by N agents, is proportional to the value of P divided by N .

Treating act-based reasons as the special case in which X is identical to P , we get the following claim about the strength of act-based reasons:

STR_{ABR} The strength, S , of agent A 's reason to do X is proportional to the value of X .

STR_{ABR} is, I take it, a plausible claim about the strength of act-based reasons. Now, this does not mean that STR_{PBR} is all or even part of the correct theory of the strength of pattern-based reasons. On the face of it, the strength of a

⁶ Smart asks: 'How can one "balance" a serious injustice, on the one hand, and hundreds of painful deaths, on the other?' 1973: 73.

⁷ See Rawls 1999a 17–18 and 42–5, Daniels 1979, and Hooker 2000: ch. 1. Rawls's response to intuitionist scepticism about theorizing about justice is admirable: 'A refutation of intuitionism consists in presenting the sort of constructive criteria that are said not to exist' 1999a: 35. Needless to say, I do not pretend to emulate his success in achieving this goal.

reason to play one's part in a pattern should reflect the importance of one's contribution in some way other than through a headcount.⁸ However, this simple theory illustrates that one test for whether we are on the right lines is to figure out the implications of our theory of the strength of pattern-based reasons for the strength of act-based reasons.

Different theories of the strength of reasons and of the function from reasons to rightness will yield markedly different theories of rightness. There is a large range of utilitarian theories here worthy of further exploration. If we suppose that few patterns are eligible, or that act-based reasons are typically much stronger than pattern-based reasons, we will get a theory close to Act Utilitarianism in its implications. If we adopt a more liberal account of eligibility and we believe that pattern-based reasons are typically stronger than act-based reasons, we will get a theory much closer to Rule Utilitarianism.

My own view on these matters, for what it's worth, lies somewhere in the middle. Here I merely report regularities in my considered judgements, rather than offering an argument or a theory. I think that act-based reasons are often decisive, but that there are some moral constraints, including those provided by moral rights. These constraints make it sometimes wrong to maximize the good. But they can also be overridden in cases where violating them is necessary to avoid great harm. As Ross claimed, it would be wrong to break a promise if breaking it would have only slightly better consequences.⁹ On the other hand, it would be wrong to keep a promise if breaking it would have much better consequences, and it would be right to break even the most serious moral constraint in order to prevent a disaster.

9.2 Cluelessness

According to Act Utilitarianism the reasons for or against an action, and its rightness, depend on the value of its consequences and on the value of the consequences of all relevant alternatives to it. According to the more complex theory we have been exploring, reasons and rightness depend on all of these facts and on further facts about the value of the consequences of all eligible patterns of action. Since we are concerned with non-perspectival reasons and rightness, these are facts about how the future will or would in fact turn out, not about what anyone would or could reasonably expect.

⁸ Compare Parfit's discussion of what he calls the Share of the Total view. Parfit 1987: 67–70.

⁹ Ross 2002/1930: 34–5.

It is clear that creatures like us could never know all of these facts. For any given token action, we would first need to know all of its consequences—or, more precisely, all of its reason-giving consequences. Though we can make some predictions, only some of the consequences of any token action could reasonably be predicted by creatures like us. In James Lenman's useful terminology, only some of them are 'visible'. Visible consequences are those we can reasonably predict, including those we are uncertain about; invisible consequences are those that cannot reasonably be predicted even by 'an ideally conscientious agent, where this is nonetheless a *human* agent and not some Laplacean fantasy'.¹⁰ In general, we have no reason to think that all of the consequences of an action are visible. The consequences of an action do not peter out like ripples on a pond, but keep going.¹¹ Further, we have no good reason to think that the invisible consequences of any token action will resemble the visible ones. What starts out having good effects can have bad effects later on, and vice versa.

Even if we knew all of the consequences of a token action, we would not know other crucial things about it. It is a further step to know the value of those consequences, which is a matter (for utilitarians) of knowing how much well-being they contain. Not even those who are optimistic about measuring the well-being of existing people think we can measure the well-being of future people. Nor do we know what all the relevant alternative actions are. In Dale Miller's example:

When deciding what to fix for breakfast this morning it did not occur to me to bolt out my front door, run two blocks east, and burst into the corner house, yet this would have prevented a great deal of disutility if the occupant of that house were choking and needed someone to perform the Heimlich Maneuver.¹²

And, of course, even if we did know all of the relevant alternatives, we would still need to know all of their consequences, and how much well-being they contain, and all of the eligible patterns, and their consequences, and how much well-being they contain.

¹⁰ Lenman 2000: 363, emphasis in the original.

¹¹ Moore 1993/1903: 201–3 notices the importance of assuming that they do peter out, while Smart 1973: 33 endorses this assumption. Lenman 2000: 350–1 points out that we have good reason to reject the assumption at least with respect to 'the very large numbers of actions that are identity-affecting' (351).

¹² Miller 2003: 50. See also Frazier 1994: 45, Howard-Snyder 1997, and Wiland 2005.

There is certainly room for debate about whether we can form justified beliefs about these things. An important issue in that debate is whether it is rational to form beliefs using a so-called ‘principle of indifference’. Such a principle tells us (at least in some cases) to assign equal credences to mutually exclusive propositions, when we must believe one of them and we have equal evidence for each of them—including when we have no evidence for any of them. For example, if you know that an envelope contains either nothing or a million dollars, but you have no evidence bearing on which of these it contains, a principle of indifference tells you to assign equal credence to each possibility.¹³

But our present question is not about what it is reasonable or justified to believe. It is whether we are in a position to *know* all of these facts about the actual consequences of actions, patterns, and their alternatives—and it seems certain that we are not. In fact, we can know only a tiny fraction of them—so small a fraction that we are, in Lenman’s description, ‘clueless’ about the things that, according to our theory, determine what reasons for action there are, and which token actions are right.¹⁴

The explanation for this is ultimately very simple. Our theory adopts non-perspectival concepts of reasons and rightness, and claims that reasons and rightness depend on facts about the far future in the actual world, as well as on facts about many different possible worlds, including their far futures. Because reasons and rightness, according to this theory, depend on these distant facts, and not on what it is reasonable to believe about them from any perspective, they are not only non-perspectival, but radically detached from anyone’s perspective. We may be able to form reasonable beliefs about reasons and rightness in some real cases, but we cannot know, in any real case, any but a tiny proportion of the reasons for action someone has, nor which of her actions is right.¹⁵ We cannot know these things even ‘retrospectively’. We could not look back at all of the consequences of an action unless we were standing at the end of history—and even then we could not know what all of the consequences of all of the relevant alternatives to it would have been.

¹³ Lenman 2000: 353–9 casts doubt on use of this principle in forming beliefs about the consequences of actions. Mason 2004 and Greaves 2016 argue that in at least some cases it is reasonable.

¹⁴ Lenman 2000.

¹⁵ I have claimed that reasons are associated with specific goods, such as *Jack’s pleasure now*. This implies that we can know about some reasons, since we can know for example that, if the agent were to perform *X* this would cause Jack some pleasure. However, most reason-giving consequences of token actions are not visible in this way. So, we cannot know most reasons, and we cannot know which token actions are right. I am grateful to Neil Sinclair for discussion here.

9.3 Good Decision Procedures

Since we cannot know which token actions are right (in the non-perspectival sense), we cannot use such knowledge to take decisions. As we noted in Chapter 3, we must also reject Moore's proposal to define good decision-making in terms of justified beliefs about rightness, since that proposal founders on Jackson Cases. We need to give some other account of good decision-making. As it is standardly put, we need to give some account of the appropriate utilitarian 'decision procedure'.¹⁶

This phrase, however, is ambiguous. In one sense of 'decision procedure', it refers to a rule or method that can be deliberately followed, step by step, to make a decision. I shall call decision procedures in this sense 'decision rules'. They specify steps to follow in order to make a good decision. Examples include 'estimate the consequences of all of your options, and choose the option with greatest expected value', 'perform the action you feel intuitively to be best', 'determine which of your options is legally permitted, and choose any of them', and 'consult someone wise, and follow their advice'.

There is a different sense of 'decision procedure', however, which we must not confuse with the first. In a broader sense, a decision procedure is any psychological process that results in an action, whether or not it includes deliberately following a decision rule, such as one of those just mentioned. It is plausible that most actions are not preceded by conscious deliberation. They are still the result of psychological processes, however, and the broader sense of 'decision procedure' includes those processes. The psychological processes underpinning exercises of skill, habits, and unreflective responses to circumstances constitute parts of a person's decision procedure in this broad sense, along with the processes involved in conscious deliberation. I shall call decision procedures in this broader sense 'decision characteristics'.

Whether we are considering decision rules or decision characteristics, we need to take account of a further distinction. This is the distinction between something's being good or bad, and its being good or bad *qua* decision procedure. The most obvious way for consequentialists to evaluate a decision procedure is, simply, in terms of all of its effects. But this is like evaluating a scientific method in terms of all of its effects. Among the effects of a new scientific method may be fame and riches for its inventor, but we do

¹⁶ See Bales 1971.

not think that these sorts of effect increase its value as a scientific method. If we want to evaluate it qua scientific method, we need to focus attention on a subset of its effects—namely the ones relevant to its function of generating scientific knowledge.

Similarly, if we want to evaluate decision procedures qua decision procedures, we need to focus on a subset of their effects, namely the ones relevant to the functions of decision procedures. I assume that the function of decision procedures is to issue in right actions by the decision-maker. If so, we might not count some good effects of decision procedures as relevant to their value as decision procedures. For example, suppose that following some procedure causes me to violate lots of rights, but in a very minor way. If others find these minor wrongs on my part highly amusing, that will mitigate the total effects of this procedure. But we might not count that amusement as tending to make the decision procedure better as a decision procedure, since it is a good consequence of the procedure that does not make the actions it causes right. Of course, Act Utilitarians will not recognize a meaningful distinction between the good effects of a decision procedure and its tendency to cause right actions, since any effects of a decision procedure that contribute to its goodness will also contribute to the rightness of the actions it leads to, on their view. But we can draw the distinction meaningfully, since on our view an action can be wrong despite being better than any other alternative.

For some purposes it may be interesting to ask, simply, which decision procedures are best. This is a way of asking which psychological characteristics are best. But I will assume that our focus is on which decision procedures are best as decision procedures. This means that the relevant test is whether they tend to cause the agent to act rightly—since this is, I assume, the function of decision procedures.

Now, the question immediately arises of whether we are as clueless about decision procedures as we are about the rightness of token actions. If we cannot know which token actions are right, it may seem to follow that we cannot know which decision procedures tend to cause agents to act rightly. If that is correct, we may be able to specify what it would take for a decision procedure to be good, but we will not be able to identify which ones have this property. In Robert Frazier's helpful way of putting it, we will not be able to 'validate' any decision procedure.¹⁷

¹⁷ Frazier 1994: 45–9. Frazier argues that no decision procedure for Act Utilitarianism can be both validated and something we could use.

However, this follows straightforwardly only if determining which token actions are right is necessary for validating decision procedures. If there is some other way to validate decision procedures—that is, to identify which ones tend to cause the agent to act rightly—then we may be able to validate them even though we are clueless about the rightness of token actions.

Happily, there does seem to be a way to do this. The key point is that each decision procedure's value depends on its tendency to cause right action, not on whether it causes right action on any single occasion. Any claim about the value of a given decision procedure will therefore be a generalization across multiple uses of it. This helps because, while we cannot assume that the invisible consequences of a token action will resemble its visible consequences, it is much more reasonable to assume that the invisible consequences of actions *of a certain type* will resemble the visible consequences of actions *of that type*. So far as its visible consequences go, a token act of promise-keeping may be very good; yet, its invisible consequences may be much worse, making its total consequences bad overall. But unless there is some reason to expect that this sort of thing happens systematically with promise-keeping, we can form reasonable beliefs about the total consequences of acts of promise-keeping, as a type, on the basis of the visible consequences of the token acts of promise-keeping that we have observed. The visible consequences are a reasonable sample of the total consequences when we generalize across cases, even though they are not a reasonable sample of the total consequences of any token act.¹⁸

Note that two things have happened here. We have shifted attention from token actions to types of action, and we have shifted attention from knowledge to reasonable belief. The claim in Section 9.2 was that we cannot know the rightness of any token action. The claim here is that we can have reasonable beliefs about the tendencies of decision procedures to lead to right actions. Since the value of decision procedures depends on many instances of use, this is a reasonable belief about one or more types of action—the type or types of action selected by the decision procedure.

Hence cluelessness about token actions does not entail that we cannot validate decision procedures. The extent to which we can validate them is no

¹⁸ My remarks here are indebted to Burch-Brown 2014. She also helpfully points out some limitations of these considerations. For example: 'Oedipus was not undermined fundamentally by a failure of prediction' (11). Lenman 2000: 360 briefly considers the idea that we can evaluate decision procedures by their consequences, but dismisses it without considering the significance of the type/token distinction. For discussion of Rule Consequentialism and epistemic problems see Hooker 2014: 183–4.

doubt limited, and our beliefs about them are likely to be flawed in various ways. But it seems that Mill was right:

...mankind must by this time have acquired positive beliefs as to the effects of some actions on their happiness; and the beliefs which have thus come down are the rules of morality for the multitude, and for the philosopher until he has succeeded in finding better...that mankind have still much to learn as to the effects of actions on the general happiness, I admit, or rather, earnestly maintain.¹⁹

Moreover, we can study decision-making scientifically, and we can hope to improve our beliefs about decision procedures substantially.²⁰

Consider, then, what it is reasonable to believe about the value of decision procedures in the sense of decision rules. First, it seems clear that different kinds of decision problem call for different decision rules. If speed is of the essence, decision rules requiring extensive calculation will not fare well. If there are high stakes, going with what seems intuitively best may not be the best approach. If others have relevant expertise, the best decision rule may recommend following their advice. If you are playing rock-paper-scissors, randomizing will be a good decision rule.

So we might expect different decision rules to be valuable in different cases. None of them will be best in every case, if only because following a decision rule is itself incompatible with making a good decision in some kinds of case. If you are up on the dance floor you should not employ any rule at all to decide what movement to make next. Importantly, this implies that there cannot be a good all-purpose meta-decision rule, which tells you which decision rule to employ for each kind of decision problem. Sometimes it will be good to use one decision rule, and sometimes it will be good to use another, and sometimes it will be good not to use a decision rule at all. We

¹⁹ Mill 1998/1861: 70. Compare Crisp: 'Just as the notion that one should keep children away from dangerous machinery is passed down from one generation to another—in this case, not explicitly, but as part of a whole network of injunctions, prohibitions and advice concerned with safety—so much other wisdom distilled over the ages is taught to us in such a way as to form moral or prudential dispositions. This wisdom has already been tested' 1992: 148.

²⁰ One domain in which this has been studied extensively is medical decision-making: see Bornstein and Emler 2001 for a review. There is also an extensive literature on the effects of heuristics and biases on decision-making more generally. Kahneman 2011 summarizes research in the tradition that conceives of heuristics and biases as ways of falling short of rationality, while Gigerenzer 2008 articulates an 'ecological' conception of rationality, according to which discoveries about heuristics and biases are bad news for our previous conception of rationality, not for rationality itself. Tetlock and Gardner 2016 describes fascinating empirical work on the characteristics of good predictors.

can hope to develop dispositions to use decision rules appropriately, but we will be going wrong some of the time if we always select decision rules using a meta-decision rule.²¹

Now, we should check that this way of thinking about good decisions has acceptable implications about the difficult 'Jackson Cases' that we discussed in Chapter 3. Recall Jackson's case of a doctor who must decide which medicine to prescribe to a patient with a relatively minor complaint. The doctor knows that treatment *A* would provide a partial cure, and she knows that one of *B* and *C* would provide a total cure, while the other would kill the patient. However, she has no way to find out which of *B* and *C* would cure the patient and which would kill him.²² According to our theory of rightness, either prescribing *B* is right or prescribing *C* is right; we do not know which is right, but we do know that prescribing *A* is wrong. However, I assume that we want an account of good decisions to imply that the best decision for the doctor is to prescribe *A*, since anything else is too risky. The question is whether the account I have sketched of good decision rules can generate this result.

At first it appears not. Following any rule that recommends prescribing *A* will lead with certainty to a wrong action, while following any rule that recommends prescribing *B* or *C* has some chance of leading to a right action. Thus it appears that a rule that recommends prescribing *A* must be worse, on the view we have been considering, than some alternative that recommends prescribing *B* or *C*. This is the wrong result.

The problem arises if we simply count wrong actions. Then the minor wrong of failing to cure the symptoms completely is counted as of equal importance as the very significant wrong of killing the patient. Obviously, we need also to take account of the seriousness of these wrongs in some way.²³

Quite how to do that is a very large topic in itself. It would be easier if the rightness of an action correlated strictly with its goodness, because we could then say that the extent to which a decision rule leads to right actions is equal to its expected value. But our theory of rightness does not allow this. Still, I will assume that we can make some reasonable judgements about overall wrongness that take account of the seriousness as well as the number of wrong actions. We might judge, for example, that a rule that leads to one murder causes more seriousness-weighted wrongness of actions than a rule that leads to three promises being broken. Similarly, we can get the right

²¹ Compare Crisp 1992: 158–9.

²² See Jackson 1991: 462–3. We considered this case in Chapter 3, Section 3.3.

²³ Bykvist 2009 makes a similar claim.

answer in Jackson's case if we claim that a rule that leads to n deaths leads to more seriousness-weighted wrongness of actions than a rule that leads to $2n$ failures to cure completely.

Finally, then, consider the other sense of decision procedures: decision characteristics. To specify a good decision procedure in this sense is to specify psychological features of persons that tend to cause right actions—where that is measured, as we have just seen, not only by the number of wrong actions avoided but by their seriousness.

One relevant decision characteristic is being disposed to use good decision rules, and to apply them well, when appropriate. A good decision-maker may act spontaneously or intuitively much of the time, but will sometimes slow down and deliberate carefully. However, good decision characteristics may include things other than dispositions to use decision rules appropriately. For example, attending to the right things must play a role. Noticing relevant features of the environment, being disposed to spot details when helpful and to ignore them when not, will surely make a difference to the quality of decisions. Having the right kind and degree of determination or having acute senses may also contribute to having good decision characteristics. Any psychological feature that contributes to causing the agent to act rightly may be relevant.

As this may suggest, which decision characteristics are good is likely to be highly context-dependent. It will depend, first, on the environment in which the agent is situated.²⁴ Having a strong startle reflex may be a good thing in the jungle, for example, but not in an operating theatre or plane cockpit.²⁵ Second, whether a characteristic is good is likely to depend on the agent's other characteristics. Julia Driver gives the following nice example: 'Imagine a malicious klutz... someone who wanted, and intended, to cause only harm to others but because of her clumsiness ended up, on balance, doing more good for them than bad.'²⁶ Klutziness is a good decision characteristic in this person, but only because it is paired with malice. In this way, decision characteristics may be a highly individual, situation-specific, and Heath Robinson affair. A good decision characteristic need not be admirable or virtuous.

²⁴ Gigerenzer 2008 emphasizes the importance of environmental context to rationality generally. 'Human factors' or ergonomics researchers take a complementary approach to studying decision-making: see for example Endsley 1995.

²⁵ Martin et al. 2015 examines the bad effects of startle responses on pilot performance. The authors note that '[i]n the worst case scenario, a pilot might never make sense of ambiguous cues and his or her brain might be completely hijacked by task-irrelevant, anxious thoughts, at the expense of constructive analysis' 2015: 100.

²⁶ Driver 2001: 66.

9.4 Praiseworthiness

The topic of praise and blame, and the appropriateness of each, is a very large one. Much ethical thought and action is a matter of holding people responsible, evaluating their actions, motives, or characters, and distinguishing appropriate from inappropriate reactive attitudes—including praise, blame, admiration, anger, contempt, pride, and shame—towards them. We cannot hope to discuss all of these topics adequately here. Instead, our aim will be to show that there is no barrier in principle to a plausible utilitarian account of them. This is not a trivial aim, because it can appear that utilitarianism is simply unable to explain them.

The central problem is to see what form a utilitarian account of praiseworthiness and blameworthiness—and of the appropriateness of reactive attitudes generally—could take. It can seem as though there are just two alternatives, both of which are plainly unsatisfactory.

The first alternative is to claim that praise (say) is appropriate just in case it is directed towards a good act, character trait, or motive. Thus we might say that the reason why courage is praiseworthy is that being courageous is a good trait. We could then try to explain the value of courage in terms of its typical effects on well-being. Call this the ‘valuable trait’ theory of praiseworthiness.

One good feature of this theory is that it clearly enables us to say that praiseworthy traits are valuable without departing from welfarism. We typically do think that praiseworthy traits are good, and blameworthy ones are bad; and utilitarians will want to account for those judgements without committing to any claim that is inconsistent with welfarism. The valuable trait theory is a straightforward way to do this.

However, the valuable trait theory has two serious problems. The first is that, as we just saw in discussing good decision characteristics, what makes a trait good is likely to be highly context-dependent: it will depend on the circumstances, and on the other features of the person concerned. In Driver’s example, klutziness turns out to be a good feature, as does malice, when each is paired with the other. Yet though each may be good, neither is admirable or praiseworthy. To take another example, we may not wish to claim that competitiveness is an admirable trait, even if in some circumstances it has good effects.²⁷ The first problem, then, is that a trait’s being valuable does not seem sufficient for its being praiseworthy or admirable.

²⁷ This example is due to Robert Adams: see Adams 2006: ch. 4. There he is criticizing valuable trait theories of virtue.

The second problem is that there seems too little connection, on this account, between some trait's being praiseworthy or blameworthy and its being right to praise or blame it. Now, we do not usually take it that something's being praiseworthy entails that it is right to praise it, or that something's being blameworthy entails that it is right to blame it. But intuitively there should be some connection, or reasonably strong correlation, between these things.²⁸ On the valuable trait theory, there may be very little correlation. It is perfectly consistent with that theory that praising the praiseworthy, and blaming the blameworthy, is almost always wrong.

Now consider the second obvious way to construct a utilitarian account of praiseworthiness and blameworthiness. Rather than focus on the value of the traits, we could focus on the rightness of the act of praise or blame. Sidgwick wrote:

From a Utilitarian point of view... we must mean by calling a quality 'deserving of praise,' that it is expedient to praise it, with a view to its future production: accordingly, in distributing our praise of human qualities, on utilitarian principles, we have to consider primarily not the usefulness of the quality, but the usefulness of the praise.²⁹

As Sidgwick puts it, the issue is how to 'distribute' praise, and the utilitarian criterion for that is the rightness of acts of praising. Thus a trait or attribute is praiseworthy if and only if and because it is right to praise it. We can call this the 'rightness of praising' theory of praiseworthiness.³⁰

It is no surprise, according to this theory, that we expect a fairly strong correlation between something's being praiseworthy or blameworthy, and its being right to praise or blame it. This is no surprise because rightness of praising or blaming *entails* praiseworthiness or blameworthiness. But now this connection seems too strong. For example, it might be right to praise an angry tyrant in order to placate him, but we do not think that he is worthy of praise. Or, as Tim Mulgan writes:

... it may often be expedient to blame a perfectly innocent person, perhaps in order to set an example to others, but it would be odd to say that such a person was blameworthy. She doesn't deserve to be blamed any more than

²⁸ Arpaly 2000: 55–6 makes a similar point. I am indebted to her excellent discussion.

²⁹ Sidgwick 1907: 428.

³⁰ Smart seems to have taken a similar view: see Smart 1973: 53. For discussion of Smart's view, see Arpaly 2000.

an innocent person killed in order to appease a hostile mob deserves to be killed.³¹

As these examples show, we want to identify a kind of ‘appropriateness’ of praising or blaming that is logically separate from the rightness of acts of praising or blaming. This separation would allow for cases in which it is right to praise or blame but not appropriate to do so, or vice versa. Appropriateness is one contributing factor to rightness, but it seems possible for it to be outweighed by other factors, such as the tyrant’s anger and power.

The question, then, is how to account for the appropriateness of praise or blame, in such a way that it is neither simply a function of the value of the relevant trait nor simply a function of the rightness of praising or blaming. In the remainder of this section I will outline one way to do that, which follows naturally from the account of reasons proposed in this book.

Suppose that there is some practice *P* of praising and blaming, and that this practice promotes well-being. How might it do that? One mechanism might be that it aids coordination, by communicating information about relevant norms. Behaviour in accordance with those norms is routinely praised by participants in the practice, while behaviour that flouts those norms is routinely blamed. A second mechanism is that the praising and blaming might encourage people to emulate the behaviour praised, and to seek to avoid emulating the behaviour blamed.³² Note that the success of both mechanisms relies on praising and blaming being somewhat systematic; if it is too chaotic, it will not serve to communicate the norms or to encourage emulation. Thus it is plausible that a certain amount of praising and blaming, conducted in the right sort of way, can serve to inform people about recognized moral rules and to encourage people to conform with them. If the norms themselves favour actions that typically promote well-being, a practice of praising and blaming may well promote well-being.³³

³¹ Mulgan 2000: 51. As he remarks, ‘[s]he’s just in the wrong place in the wrong thought experiment’.

³² Holroyd 2007 points to evidence that if a practice of praising and blaming functions as a system of reward and sanctions, rather than a system for communicating norms that agents can internalize, it can sap rather than increase motivation to act well. She argues that this undermines consequentialist views of praise and blame (270–1), but it is not clear why it does not instead suggest that good practices of praise and blame will tend to function more as communicative systems than as systems of reward and punishment.

³³ Braithwaite 1989 claims that ‘reintegrative’ practices of shaming—roughly, those that blame persons for their actions rather than their characters, and that include processes to enable reintegration into the moral community, rather than shunning—are effective in reducing crime. See also the fascinating discussion of the ‘economy of esteem’ in Brennan and Pettit 2004.

Let us assume that, where such a practice exists, it is both valuable and eligible. According to our theory of reasons, this means that individuals have pattern-based reasons to participate in the practice. In other words, they have reasons to play their part in the activity of routinely praising behaviour that conforms with relevant norms, and of routinely blaming behaviour that flouts them. This activity is quite systematic, and the behaviour which is praised is often emulable, while the behaviour which is blamed is avoidable.

Now, it follows that, where such a practice exists, there would be, in general, reasons to praise behaviour that conforms to the norms, and reasons to blame behaviour that flouts them. In this sense, the norms would identify behaviour as appropriate to praise or blame. Since the practice as a whole promotes well-being, the norms must be in good order, from a utilitarian point of view. The behaviour they prescribe must be, in general, valuable and emulable, and the behaviour they proscribe must be, in general, harmful and avoidable.

But of course our theory of reasons is pluralist, so we cannot infer from the appropriateness of praise and blame that it is right to praise or blame. Nor can we infer from the inappropriateness of praise and blame that it is wrong to praise or blame. We have to take into account act-based reasons for praising and blaming, too. For example, if we are faced with a powerful tyrant, the fact that his behaviour flouts the norms does not entail that it is wrong to praise him. It means that it is inappropriate to praise him, in the sense that his behaviour is not praiseworthy. But placating him may be more important than appropriateness. It could be right to praise him, even though it is inappropriate to do so.

This is one way in which utilitarians can distinguish between the rightness of praising (or blaming), on one hand, and praiseworthiness (or blameworthiness), on the other, without completely dissociating them. The account I just outlined implies that in ordinary, good, cases of praise and blame, it will be right to praise or blame only if the target quality is praiseworthy (or blameworthy). But it also allows these things to come apart in unusual cases, such as with praising the angry tyrant or Mulgan's case of blaming the innocent. Note that these conclusions are based on understanding the relationship between the concepts involved, by considering their application to hypothetical cases. They do not depend on knowledge of reasons or rightness in real cases.

It is also important to note that, on the utilitarian view, a quality's being praiseworthy is not a distinct kind of value. If it were, it would violate utilitarianism's commitment to welfarism. Instead, the norms of

appropriateness that explain praiseworthiness and blameworthiness are constitutive of well-being-promoting practices. In that respect, they are akin to the rules of chess. It is inappropriate to praise someone who is not praiseworthy in roughly the same way that it is inappropriate to move a rook diagonally.³⁴

One awkward implication of this account of praiseworthiness is worth noting. If a practice of praising is eligible only if it is actual, then a quality that is not the object of a well-being-promoting practice of praise cannot be praiseworthy.³⁵ It may seem odd to claim that a quality cannot be praiseworthy unless it is praised. It is, I acknowledge, an unusual feature of this combination of views. But it follows from the respect in which the appropriateness of praise (and blame) is akin to the permissibility of a move in chess, together with this view of eligibility: if the practice did not exist, there would be nothing to ground the appropriateness or the permissibility.

This is, I take it, a revisionary feature of this view. One way to avoid it would be to adopt a different view of the eligibility of patterns. But perhaps we can learn to live with it. One thing that may help us to do that is noting that if we adopt this combination of views we can still make counterfactual claims about praiseworthiness. Suppose, for example, that a certain quality, *Q*, is highly effective in causing right actions, that it is highly emulable, and that it is currently unrecognized by all practices of praising. In fact, only a few psychologists have detected *Q* and its remarkable effects. We might then say that, though this quality is not currently being praised, it would be good if practices of praise recognized it and started to praise it. Though it is not currently praiseworthy, it ought to be praised, and if it were, it would be praiseworthy. For some purposes, perhaps, counterfactual praiseworthiness can substitute adequately for actual praiseworthiness.

9.5 Virtue

The concepts of virtue and of vice are more narrowly focused than the concepts of praiseworthiness and blameworthiness. Actions, motives, and character traits can all be praiseworthy or blameworthy in their own right.

³⁴ A significant difference is that the rules of chess are codified, whereas the norms governing well-being-promoting practices of praise and blame are not. Of course, there are many other differences, not least in the emotional and moral significance of the reactive attitudes.

³⁵ This is roughly the view of eligibility defended in Chapter 5. I claimed there that patterns other than established practices could be eligible, but here I ignore that complication.

In the case of virtue, character traits are central. Though we speak of virtuous actions and motives, we do so by courtesy of their relationship to the primary objects of virtue, which are character traits.³⁶

The standard way for consequentialists, including utilitarians, to evaluate character traits is in terms of their goodness. Thus the most obvious way for utilitarians to construct a theory of virtue is to characterize virtues as good character traits, and vices as bad ones.³⁷ However, the simplest versions of this approach run straight into the problem we encountered for the valuable trait theory of praiseworthiness, that the fact that a characteristic is valuable does not seem to be sufficient for it to be admirable. In Driver's example, the value of malice or klutziness does not suffice for either to be admirable.

Driver responds to this problem by defining a virtue as 'a character trait that produces more good (in the actual world) than not *systematically*'.³⁸ Since neither malice nor klutziness has systematically good effects, they are not virtues. This certainly seems to be an improvement over the simpler account. By requiring that good effects are produced systematically, Driver's proposal enables us to set aside, as irrelevant to virtue, those effects of a trait that have highly idiosyncratic causes. That certainly helps to explain why some good traits do not count as virtues. But it does not seem to go far enough. Some traits have systematically good effects because of stable features of the agent's background circumstances, yet do not seem to be virtues.

Driver herself offers an excellent example to make this point:

Imagine a society that has evolved differently from human society. These creatures, Mutors, have evolved in an extremely harsh environment and have developed unusual strategies for survival. It happens to be the case that for them, beating one's child severely when it is exactly 5.57 years old actually increases the life expectancy of the child by 50 percent. The child is upset by the beating, but this feeling goes away in time. It is also the case that the only way a Mutor could ever bring himself to so treat a child is to develop an intense pleasure in doing so. So some Mutors have a special

³⁶ Some doubt that there are character traits: see for example Doris 1998 and Harman 2000: ch. 10. For responses to this 'situationist' view, see Adams 2006: chs. 8–9, Sreenivasan 2002, and Webber 2006.

³⁷ See Bentham 1838–43b: ch. 11, Sidgwick 1907: 424, Shaw 1999: 255–61, and Bradley 2005. Hooker 2000: 36 outlines a Rule Consequentialist conception of virtue.

³⁸ Driver 2001: 82, emphasis in the original.

trait—they intensely desire to beat children who are exactly 5.57 years old. That it is good for the child is irrelevant to them.³⁹

The Mutors' environment is such that this trait has systematically good effects, and so on Driver's account it is a virtue, in that context. Yet this is hard to accept. The problem is not that what is good in this harsh context would not be good in our more fortunate circumstances. It is, after all, plausible that what constitutes a virtue depends on background circumstances. The problem is that, even in the circumstances described, the desire to beat children of a certain age for pleasure is in no way admirable. If being admirable is necessary for virtue, systematically good effects do not suffice to make a trait a virtue.⁴⁰

Robert Adams makes a similar point when he claims that, intuitively, virtues not only usually have good effects but are also 'excellences' of character. Consequentialist theories, he claims, can explain the first feature but not the second; whereas his own theory—that virtues are 'excellences in being for the good'—can explain both. Those who are excellent in being for the good will be good at promoting the good, and so will usually have good effects. But consequentialists will have to treat competitiveness, for example, as a virtue so long as it has good consequences—even though it is not an excellent orientation to goodness and is not, intuitively, virtuous.⁴¹

By combining our accounts of good decision characteristics and of praiseworthiness, we can outline an alternative consequentialist account of virtue that may be able to avoid these objections. Intuitively, being (practically) virtuous entails being a good decision-maker. Thus, we might expect virtuous character traits to be good decision characteristics. Furthermore, being virtuous is, intuitively, sufficient to be praiseworthy and to avoid blameworthiness. So we can consider the following hypothesis: a trait is virtuous if and only if, and because, it is a good decision characteristic (it generally leads to right decisions) *and* it is praiseworthy (it is picked out as praiseworthy by the norms of a good practice of praising and blaming). Let us briefly consider some of the features of virtues, according to this hypothesis.

Not all good decision characteristics are character traits, and of those that are, many may not be praiseworthy. They may fail to be praiseworthy because their value depends on being paired with other non-admirable

³⁹ Driver 2001: 55.

⁴⁰ See Bradley 2005 for further discussion of Driver's view.

⁴¹ Adams 2006: ch. 4. Compare Hurka 2001: 20.

traits, or because they are not emulable, and so a practice of praising them fails to promote well-being. Having an extraordinary natural facility for mental arithmetic, for example, may be a character trait and it may tend to cause right actions, since calculating probabilities can help with that. But even if it is a good character trait, and a good decision characteristic, it may not be praiseworthy, because it is not emulable. In contrast, the character trait of trying to develop one's own abilities does seem to be praiseworthy, since a practice of praising it, we may suppose, will tend to promote well-being.

If a character trait is both a good decision characteristic and praiseworthy, it will be something that tends to cause right actions and which is such that a practice of praising it promotes well-being. Any such trait will have a good claim to be called a virtue.⁴² It will be appropriate to praise this trait, and in that sense it will be both admirable and an excellence of character. This consequentialist conception of virtues will thus be better placed to respond to Adams' point. The issue will be not whether consequentialists can coherently claim that virtues are excellences of character—but instead whether they have the correct understanding of what constitutes excellence. For the utilitarian as for Adams, loving the good and hating the bad will be virtues. The difference is that Adams treats these characteristics as both excellent and noninstrumentally good, while utilitarians will treat them as excellent because instrumentally good.⁴³

Given our outline account of rightness, we can expect a strong disposition to respect rights to be a virtue in this sense. Having this disposition to act justly will tend to cause right action, and a practice of praising it will promote well-being (on the assumption which we made in Chapter 6, that the practice of respect for rights itself promotes well-being).⁴⁴ Having this disposition will therefore be an excellence of character, in the sense that it is admirable, on the present account. Other examples of virtues, we may conjecture, include loyalty to friends and loved ones, open-mindedness,

⁴² Whether the Mutors' desire to beat children who are exactly 5.57 years old is a virtue, on this account, depends on whether a practice of praising it would have good effects. There is some reason to doubt this, since if their psychology is at all like ours, a practice of praising this trait would be likely to have very harmful collateral consequences. But if it would have overall good consequences it might turn out to be a virtue. Given how difficult it is to imagine such a case clearly, it is hard to assess the significance of this.

⁴³ See McElwee 2015 for further discussion of possible consequentialist theories of virtue.

⁴⁴ It is easier to explain why justice is a virtue if we define good decision characteristics as those which tend to cause right action, rather than as those which are valuable. See Adams 2006: 59.

calmness under pressure, and beneficence. It is plausible to think that each of these traits tends to cause right action, and that a practice of praising them promotes well-being. If this is correct, we will have reason to admire them, and it will be appropriate to praise them.

Similarly, any trait that is both a bad decision characteristic and blame-worthy will be a vice, on this account. Any such trait will tend to cause wrong actions, and it will be such that a practice of blaming it promotes well-being. It will be a defect in character, and deplorable.

9.6 Conclusion

The non-perspectival account of reasons and rightness forces us to reconfigure our picture of ethical life. If rightness and reasons are, as I have claimed, non-perspectival matters, and they depend on the total consequences of actions, then agents are, quite generally, unable to detect them or to use them as guides to action. Trying to determine which option is right, or what one's reasons are, is not a good way to take decisions. About these matters, agents are almost completely in the dark.⁴⁵ We cannot characterize good decision-making as being guided by judgements about reasons or rightness. Nor can we characterize blameworthiness as essentially involving wrongness.

However, an alternative picture is possible. I have tried to show that utilitarians (and, more generally, consequentialists) who give a non-perspectival account of reasons and rightness can give a plausible account of good decision-making, of praise and blame, and of virtue. On this picture, being virtuous consists in having character traits which tend to cause right action and which are praiseworthy. Faced with a difficult decision, one good piece of advice is, therefore, 'act as the virtuous would act'. Faced with the question of what kind of person to try to be, a good piece of advice is 'become more virtuous'.⁴⁶ Since virtues are generally emulable, we can hope

⁴⁵ C. D. Broad wrote: 'We can no more learn to act rightly by appealing to the ethical theory of right action that we can play golf well by appealing to the mathematical theory of the flight of the golf-ball. The interest of ethics is thus almost wholly theoretical, as is the interest of the mathematical theory of golf or of billiards. And yet it may have a certain slight practical application. It may lead us to look out for certain systematic faults which we should not otherwise have suspected; and, once we are on the look out for them, we may learn to correct them' Broad 1930: 285.

⁴⁶ Crisp 1992: 159 claims that '[u]tilitarian moral agency requires the sort of insight into what makes life worth living that can be gained only by forswearing utilitarianism as a decision procedure. The moral life is the life of virtue, in which entrenched dispositions of concern for others, and a sensitivity to their needs and wants, supply the foundation for practical reasoning'.

to follow this advice. Since which traits are virtues is open to empirical investigation, we can also hope to learn more about what, in practice, becoming more virtuous would involve.

This alternative picture does not require us to introduce perspectival concepts of reasons or rightness. We can specify good decision rules, good decision characteristics, praiseworthy features, and virtuous traits without appealing to the idea of perspectival reasons or to the idea of perspectival rightness. We do not need to go beyond the simple picture discussed in Chapter 3, according to which there is just a non-perspectival concept of reasons and a non-perspectival account of rightness.⁴⁷

What is more, non-perspectival reasons and rightness have an essential role in the account of virtuous agency outlined here. Good decision-making is not defined as decision-making that tends to have good consequences, but as decision-making that tends to issue in right action. Virtues are good decision characteristics, so they too are defined in terms of right action. This means that the non-perspectival concepts of reasons and rightness have great significance. Their significance does not depend on their being action-guiding, but instead on their being the criterion of good decision-making.

⁴⁷ Perspectival/non-perspectival pluralists may claim that it is desirable or convenient to use perspectival concepts of reasons and rightness to characterize good decision-making, even if it is not necessary. However, doing that seems to me unhelpful, because it would suggest that norms of rationality are criteria of good decision-making. On my view, the criterion of good decision-making is the tendency to produce right action. If some kinds of irrationality generally lead to more right actions (in the non-perspectival sense), then good decision-makers will be irrational in those ways, and it may be virtuous to be irrational in those ways.

Conclusion

Utilitarianism is a family of views, or a tradition of thought, animated by a simple and beautiful idea: that what ultimately matters is that people have good lives, and that each person's well-being matters equally.¹ Utilitarians seek to explain the component parts of ethics—reasons, rightness, justice, virtue—by elaborating this simple idea.

No other tradition of ethical thought has a more humane or attractive central idea. But, of course, the simplicity of its central idea leaves utilitarianism open to the objection that it is, in the end, too simple. Our non-philosophical ethical ideas and responses are complex and nuanced, and if philosophical theories of ethics must cohere with them to a reasonable degree, utilitarianism may seem ill equipped.

Yet the utilitarian tradition is full of suggestions about how to explain this complexity. Many of the best suggestions were made by Mill or Sidgwick, long ago. Some utilitarians stick with simple versions of the theory, and place all the explanatory weight on empirical claims. Others seek to incorporate more complexity into the theory itself, as I have done in this book.

We began by identifying six powerful objections to utilitarianism, and wondering whether a single version of utilitarianism could give plausible replies to all six. Since then, we have been exploring a version of utilitarianism with some distinctive features. Chief among these features are its non-perspectival and pluralist account of reasons, its account of moral rights and justice, and its account of virtue. Do these features enable it to give plausible replies to the six objections we identified?

The first objection was that utilitarianism has an inadequate theory of value. As I defined it, utilitarianism is committed to two claims about value: welfarism, and sum-ranking. Each of these is, of course, open to objection and is the subject of much debate. But neither claim, nor their conjunction, is obviously wrong. It is not obvious that some well-being lacks

¹ Of course, utilitarians think that the well-being of all possessors of well-being matters. We have set aside non-persons in this book, but not because they do not matter.

noninstrumental value, nor that anything other than well-being possesses it. For most candidate values, including beauty, knowledge, and autonomy, it is quite plausible that the value in question is explicable entirely in terms of its contribution to well-being—as a constituent of well-being, a cause of it, or both.

The case of substantive equality is perhaps the hardest. It is quite plausible to claim that the way in which goods are distributed matters in itself, separately from any effect of the distribution on the total amount of goods. Thus egalitarian or prioritarian objections to sum-ranking have quite a lot of plausibility. However, I argued in Chapter 7 that purely axiological versions of this objection are not compelling. If we focus on cases that we can be sure we intuitively grasp, it is not at all clear that sum-ranking is false. If we broaden our focus, and ask not about the value of outcomes but about what should be done to remedy great inequality, utilitarianism is on stronger ground, and has more theoretical resources to call upon. Utilitarians can try to develop egalitarian theories of justice or, as I did, characterize egalitarian aims as good policy.

This illustrates a more general point. When we think about theories of value, it is hard to be sure how to interpret our own convictions. We may be sure that something matters greatly, and leap to the conclusion that it has noninstrumental value. But there are many ways to articulate mattering. For example, something may matter greatly in the sense that there are weighty reasons to bring it about or protect it. Something else may matter in the sense that it is virtuous to admire it. If reasons and virtues have a complex relationship to value, as I have claimed, it does not follow from something's mattering, in either of these senses, that it is noninstrumentally good.

The second objection was that utilitarianism permits abhorrent actions—or, at any rate, actions that are clearly wrong. It is said to permit acts of torture, enslavement, organ snatching, wrongful punishment—and even failure to return library books or to keep promises. Since most people are not moral absolutists, the most pertinent objection is not that utilitarianism sometimes permits acts of these kinds, because that is also true of the views advocated by most of its critics. Instead, it is that utilitarianism permits acts of these kinds on some occasions when they are clearly wrong.

Obviously, there are two main ways for utilitarians to respond to this objection. One is to accept that their theory permits these acts on the specified occasions, but to deny that these acts are wrong. The other is to deny that their theory permits these acts on those occasions. My strategy has tended to be of the second kind. I have tried to show how utilitarians can

explain moral constraints, so that they can agree that it is wrong to violate people's rights by snatching their organs, torturing them, or punishing them when they have not committed any crime. If utilitarians can explain the existence of moral constraints of these kinds, they can explain why it is wrong to do these things unless failure to do them would result in a catastrophe.

In my view, the most promising way to explain moral constraints uses the concept of pattern-based reasons. Because pattern-based reasons are reasons to play one's own part in some valuable pattern of action, they are naturally agent-relative. This means that we can try to explain why Smith has an agent-relative reason not to torture without making implausible claims about value such as that, when it comes to Smith's reasons, Smith's torturing someone is much worse than someone else's torturing them. The disvalue of torture does not seem to be agent-relative in this way. By using the concept of pattern-based reasons, we can avoid distorting our claims about value.

Whether the explanation in terms of pattern-based reasons is successful depends, for the most part, on whether we can give a good explanation of the eligibility of patterns and of the strength of reasons. I have not presented a complete theory of these things here. By the same token, we have no good reason to think that these tasks are harder than those facing any other attempt to explain moral constraints. The incompleteness of the theory I have developed is not a reason to think that the objection from abhorrent actions is fatal to utilitarianism. If utilitarians want to try to account for moral constraints, they certainly have some theoretical options for doing so. For example, they could try to go further than I have done in developing a theory of the eligibility of patterns and the strength of reasons. Alternatively, if they are not moved by Dietz's argument they could adopt an idealizing form of Rule Utilitarianism.

The next objection was that utilitarianism is too demanding. This objection is extremely hard to assess, for several reasons. One is that, as we noted in Chapter 2, it immediately raises very difficult questions about which of our convictions, if any, we should trust. On one hand, it is very hard to believe that someone who works tirelessly for others is acting wrongly whenever she fails to maximize well-being, as Act Utilitarianism implies. On the other, those of us able to spend time debating these questions have a vested interest in concluding that morality is not as demanding as Act Utilitarianism implies. Another difficulty is that, to judge whether a moral theory is too demanding, we need to make up our minds about how demanding morality is—which is to say that we need to know, or think we

know, what morality requires. Reaching a verdict on the demandingness objection seems to require reaching a verdict about morality as a whole.

Yet there is good reason to think that, however demanding morality is, the version of utilitarianism I have outlined can account for it. This is simply because it can be either as demanding as Act Utilitarianism or as undemanding as libertarianism, or at some level in between, depending on how we settle some questions that I have left open. The most important question in this respect is the function from reasons to rightness. In Chapter 9 I pointed out that utilitarians are not bound to accept that the right action is always the one for which there is strongest reason. If they claim instead that an action is right so long as it is either the one for which there is strongest reason, or one which the agent has a moral right to perform, their theory will be much less demanding. If it assigns rights that govern much of life, it may be very undemanding indeed.

If that is correct, the version of utilitarianism I outlined can include options (or 'prerogatives') as well as moral constraints. If moral rights entail moral liberty, in Hohfeld's sense of 'liberty', then agents are morally permitted to exercise their rights in any way they see fit. Exactly how demanding that turns out to be depends on the scope of agents' moral rights. On the other hand, if we deny that moral rights entail moral liberty, we will have a theory that is roughly as demanding as Act Utilitarianism. Agents will be subject to the constraints provided by others' rights, but they will be morally required to exercise their own rights in the way for which there is strongest reason. Depending on how these questions about rights are settled, we can end up with a highly demanding or an undemanding theory. Thus we can conclude with confidence that the demandingness objection is not obviously successful against this form of utilitarianism.

This brings us to the claim that utilitarianism fails to respect the separateness of persons. Taken in one way, this is simply a denial of sum-ranking. The idea that we should respect the separateness of persons in some way that utilitarianism fails to do might boil down, ultimately, to the claim that the way in which goods are distributed across lives makes a difference to the value of outcomes independently of its effect on the total stock of goods. If that is what it means, utilitarians have a plausible reply, as I have already noted. Once we focus on cases that we are sure we can grasp intuitively, it is not at all clear that sum-ranking is false.

The idea that utilitarianism fails to respect the separateness of persons is more interesting if interpreted, instead, as a claim about what utilitarianism recommends doing to or for people. For example, if utilitarianism were to

recommend imposing great costs on someone for the sake of tiny benefits to millions of others, this might be said to fail to respect the separateness of the beneficiaries' lives, and to wrongly impose on the victim. This is the kind of case—especially where the victim is already much worse off than the beneficiaries—where utilitarianism may seem to care, wrongly, about well-being itself, rather than about people.

But it is not at all clear that utilitarianism of the sort that I have outlined would recommend doing this. Once again, the crucial factor is its incorporation of constraints. These imply that some ways of making outcomes better are morally wrong. Among those ways, quite plausibly, will be ones in which great costs are imposed on someone who is already badly off for the sake of tiny benefits to very many others.

Thus the reply to the separateness of persons objection depends on exactly how it is spelled-out. If it is a denial of sum-ranking, the reply is that the cases that are typically taken to show sum-ranking to be false do not do so. If it is an objection to imposing costs on some for the sake of giving benefits to others, the reply is that we must distinguish cases. In many cases, the right thing to say is that it is morally right to impose costs on some for the sake of benefits to others, and any theory that denies this is simply wrong. In other cases, we may wish to agree that it is wrong to impose costs on someone even when doing so makes the outcome better. But in these cases we may be able to appeal to moral constraints, including moral rights, to explain this judgement.

The fifth objection was that utilitarianism is inadequate when considered as a political philosophy. The thought is that utilitarianism entails a simple and technocratic picture of politics, in which experts apply moral principles to social or economic issues. In contrast, critics emphasize that politics involves contest between conflicting reasonable ways of seeing the world, and between conflicting interests. As a result, the legitimacy of procedures is a central concern in politics. On the face of it, utilitarianism may seem incapable of explaining the importance of legitimate procedures; relatedly, it may seem incapable of explaining the full importance of democracy.

However, I claimed in Chapter 8 that utilitarianism is capable of explaining these things. In particular, I specified a 'thin' sense of legitimacy, according to which a procedure is legitimate to the extent that people subject to it comply with its decisions for reasons other than coercion. This account treats legitimacy as a sociological fact, in a broadly Weberian way. But though legitimacy is not, on this account, a normative property, it is a sociological property with great normative significance. It constrains what

can be achieved non-coercively. Given the costs of coercion, this is an important constraint. By utilitarian lights, legitimacy in this thin sense gives rise to content-independent reasons, in accordance with our pre-theoretical understanding of the concept.

Moreover, utilitarians can give qualified support to democratic procedures. Democratic procedures fulfil epistemic functions and serve as safeguards against the accumulation of power. Utilitarians are not democrats as a matter of fundamental principle, admittedly. But in most circumstances they should support democracy, while noting and seeking to ameliorate its problems and limitations. Chief among these are its tendency to give insufficient weight to the interests of non-voters, including foreigners, future people, and non-human animals.

Thus utilitarianism, properly elaborated and understood, does not imply that the best form of politics is the application of moral principles by technocrats. It implies that the legitimacy of procedures gives rise to content-independent reasons that often affect what ought to be done. It also implies that whether a decision was taken democratically typically matters a great deal—not least because it affects what can be done legitimately. And it implies that considerations of justice constrain what ought to be done. It offers, in fact, a rich and fruitful way to understand the complexity and messiness of political life.

The final objection was that utilitarianism is committed to implausible claims about the psychology of persons, and to an unattractive picture of a life of virtue. The simplest version of this objection imagines that utilitarians favour what I called ‘pervasive calculation’: a policy of using calculation to make every decision. Of course, utilitarians do not favour this, as they have often pointed out. Pervasive calculation is not a good decision procedure, for straightforward utilitarian reasons: to the extent that it can be implemented, it will often have bad effects.

The more interesting version of the objection is that utilitarianism either lacks an account of virtuous agency, or is committed to a highly implausible account of it. One aspect of this challenge is to explain what would be a good decision procedure, by utilitarian lights. Another is to give a plausible utilitarian account of virtues and vices. I tried to answer these challenges in Chapter 9. There I argued that, though we are ‘clueless’ about the rightness of token actions, we are able to make sensible judgements about the value of different decision procedures. I also argued that we can avoid some of the objections to standard consequentialist accounts of virtue by defining virtues as character traits that are both good decision characteristics

(they typically cause the agent to act rightly) and also praiseworthy (they are picked out as appropriate for praise by the norms of a practice of praising that promotes well-being). On this account, virtues are not simply good psychological features; they are also admirable, and in that sense excellent.

I conclude that none of the six objections we identified in Chapter 2 is obviously fatal to the version of utilitarianism I have outlined. It allows us to give plausible replies to all six objections. This is not, I hasten to add, equivalent to claiming that it gets all these issues right. It is too early to tell. But we can at least say that it may be right: none of the objections obviously succeeds. Like other moral theories, it is worth taking seriously.

There were two aims in developing the particular version of utilitarianism presented in this book. One was to motivate interest in this specific theory. However, there was also a more important and larger purpose. As will have been clear, the version of utilitarianism I have outlined is incomplete in many respects, and speculative in many others. It may turn out to be on the wrong lines altogether. The larger purpose was to motivate renewed interest in the utilitarian tradition itself. For this purpose, the theory I have outlined is presented as an example. If, as I have claimed, it is not obviously defeated by the six objections we identified, then those objections cannot have shown that utilitarianism is a dead end. Given the attractiveness of utilitarianism's central idea, this should encourage others to contribute to the further development of the broader tradition.

Bibliography

- Abizadeh, A. 2008. 'Democratic Theory and Border Coercion: No Right to Unilaterally Control Your Own Borders'. *Political Theory* 36: 37–65.
- Adams, N. P. 2018. 'Institutional Legitimacy'. *The Journal of Political Philosophy* 26: 84–102.
- Adams, R. M. 1999. *Finite and Infinite Goods: A Framework for Ethics*. New York: Oxford University Press.
- Adams, R. M. 2006. *A Theory of Virtue: Excellence in Being for the Good*. Oxford: Clarendon Press.
- Albee, E. 1902. *A History of English Utilitarianism*. London: Swan Sonnenschein.
- Alexander, L. 1985. 'Another Look at Moral Blackmail'. *Philosophy Research Archives* 10: 189–96.
- Alexandrova, A. 2016. 'The Science of Well-Being'. In G. Fletcher (ed.) *The Routledge Handbook of Philosophy of Well-Being*. Abingdon: Routledge, 389–401.
- Alvarez, M. 2018. 'Reasons for Action, Acting for Reasons, and Rationality'. *Synthese* 195: 3293–310.
- Anderson, E. S. 1999. 'What Is the Point of Equality?' *Ethics* 109: 287–337.
- Anderson, E. S. 2010. 'The Fundamental Disagreement between Luck Egalitarians and Relational Egalitarians'. *Canadian Journal of Philosophy* 36: 1–23.
- Anderson, S. A. 2008. 'Of Theories of Coercion, Two Axes, and the Importance of the Coercer'. *Journal of Moral Philosophy* 5: 394–422.
- Anomaly, J. 2005. 'Nietzsche's Critique of Utilitarianism'. *Journal of Nietzsche Studies* 29: 1–14.
- Anscombe, G. E. M. 1958. 'Modern Moral Philosophy'. *Philosophy* 33: 1–19.
- Aristotle. 1976. *The Ethics of Aristotle: The Nicomachean Ethics*. Trans. J. A. K. Thomson, revised by H. Tredennick. Harmondsworth: Penguin.
- Arneson, R. J. 1989. 'Equality and Equal Opportunity for Welfare'. *Philosophical Studies* 56: 77–93.
- Arneson, R. J. 1999a. 'Human Flourishing Versus Desire Satisfaction'. *Social Philosophy and Policy* 16: 113–42.
- Arneson, R. J. 1999b. 'Equality of Opportunity for Welfare Defended and Recanted'. *The Journal of Political Philosophy* 7: 488–97.
- Arneson, R. J. 2001. 'Against Rights'. *Philosophical Issues* 11: 172–201.
- Arneson, R. J. 2003. 'Defending the Purely Instrumental Account of Democratic Legitimacy'. *The Journal of Political Philosophy* 11: 122–32.
- Arneson, R. J. 2015. 'Basic Equality: Neither Acceptable nor Rejectable'. In U. Steinhoff (ed.) *Do All Persons Have Equal Moral Worth? On 'Basic Equality' and Equal Respect and Concern*. Oxford: Oxford University Press, 30–52.
- Arpaly, N. 2000. 'Hamlet and the Utilitarians'. *Philosophical Studies* 99: 45–57.

- Ashford, E. 2000. 'Utilitarianism, Integrity, and Partiality'. *The Journal of Philosophy* 97: 421–39.
- Asimov, I. 1989. 'Franchise'. In I. Asimov and M. H. Greenberg (eds) *Election Day 2084: A Science Fiction Anthology on the Politics of the Future*. Buffalo, NY: Prometheus Books, 11–24.
- Asimov, I., and Greenberg, M. H. (eds) 1989. *Election Day 2084: A Science Fiction Anthology on the Politics of the Future*. Buffalo, NY: Prometheus Books.
- Austin, J. 1832. *The Province of Jurisprudence Determined*. London: John Murray.
- Bacharach, M. 1999. 'Interactive Team Reasoning: A Contribution to the Theory of Co-operation'. *Research in Economics* 53: 117–47.
- Bacharach, M. 2006. *Beyond Individual Choice: Teams and Frames in Game Theory*, ed. N. Gold and R. Sugden. Princeton: Princeton University Press.
- Bailey, J. W. 1997. *Utilitarianism, Institutions, and Justice*. New York: Oxford University Press.
- Bales, R. E. 1971. 'Act-Utilitarianism: Account of Right-Making Characteristics or Decision-Making Procedure?' *American Philosophical Quarterly* 8: 257–65.
- Bardsley, N., Mehta, J., Starmer, C., and Sugden, R. 2010. 'Explaining Focal Points: Cognitive Hierarchy Theory versus Team Reasoning'. *The Economic Journal* 120: 40–79.
- Barry, B. M. 1989. *Theories of Justice: A Treatise on Social Justice, Volume 1*. Berkeley: University of California Press.
- Ben-Ner, A. and Putterman, L. (eds) 1998. *Economics, Values, and Organization*. Cambridge: Cambridge University Press.
- Bentham, J. 1838–43a. *The Works of Jeremy Bentham*, ed. J. Bowring. Edinburgh: William Tait.
- Bentham, J. 1838–43b. 'An Introduction to the Principles of Morals and Legislation'. In J. Bowring (ed.) *The Works of Jeremy Bentham*, vol. 1. Edinburgh: William Tait, 1–154.
- Bentham, J. 1838–43c. 'Constitutional Code'. In J. Bowring (ed.) *The Works of Jeremy Bentham*, vol. 9. Edinburgh: William Tait, 1–647.
- Benz, M. and Stutzer, A. 2004. 'Are Voters Better informed When They Have a Larger Say in Politics?' *Public Choice* 119: 31–59.
- Bergström, L. 1966. *The Alternatives and Consequences of Actions*. Stockholm: Almqvist and Wiskell.
- Binmore, K. 1998. 'A Utilitarian Theory of Political Legitimacy'. In A. Ben-Ner and L. Putterman (eds) *Economics, Values, and Organization*. Cambridge: Cambridge University Press, 101–32.
- Bornstein, B. H., and Emler, A. C. 2001. 'Rationality in Medical Decision Making: A Review of the Literature on Doctors' Decision-making Biases'. *Journal of Evaluation in Clinical Practice*, 7: 97–107.
- Bradley, B. 2005. 'Virtue Consequentialism'. *Utilitas* 17: 282–98.
- Bradley, B. 2014. 'Objective Theories of Well-Being'. In B. Eggleston and D. E. Miller (eds) *The Cambridge Companion to Utilitarianism*. Cambridge: Cambridge University Press, 220–38.
- Bradley, B. 2016. 'Well-Being at a Time'. *Philosophic Exchange* 45: 1–12.

- Bradley, B. 2018. 'Contemporary Consequentialist Theories of Virtue'. In N. Snow (ed.) *The Oxford Handbook of Virtue*. Oxford: Oxford University Press, 398–412.
- Braithwaite, J. 1989. *Crime, Shame and Reintegration*. Cambridge: Cambridge University Press.
- Brandt, R. B. 1979. *A Theory of the Good and the Right*. Oxford: Clarendon Press.
- Brandt, R. B. 1992. *Morality, Utilitarianism, and Rights*. Cambridge: Cambridge University Press.
- Brandt, R. B. 1996. *Facts, Values, and Morality*. Cambridge: Cambridge University Press.
- Brennan, G., and Pettit, P. 2004. *The Economy of Esteem: An Essay on Civil and Political Society*. Oxford: Oxford University Press.
- Brink, D. O. 1986. 'Utilitarian Morality and the Personal Point of View'. *The Journal of Philosophy* 83: 417–38.
- Brink, D. O. 2013. *Mill's Progressive Principles*. Oxford: Clarendon Press.
- Broad, C. D. 1930. *Five Types of Ethical Theory*. London: Routledge.
- Broome, J. 1991. *Weighing Goods: Equality, Uncertainty and Time*. Oxford: Basil Blackwell.
- Broome, J. 2004. *Weighing Lives*. Oxford: Oxford University Press.
- Broome, J. 2013. *Rationality through Reasoning*. Oxford: Wiley Blackwell.
- Brown, C. 2011. 'Consequentialize This'. *Ethics* 121: 749–71.
- Bruder, M., and Tanyi, A. 2014. 'Overdemanding Consequentialism? An Experimental Approach'. *Utilitas* 26: 250–75.
- Buchanan, A. E. 2003. *Justice, Legitimacy, and Self-Determination: Moral Foundations for International Law*. Oxford: Oxford University Press.
- Burch-Brown, J. M. 2014. 'Clues for Consequentialists'. *Utilitas* 26: 105–19.
- Butler, J. 2006a. *The Works of Bishop Butler*, ed. D. E. White. Rochester: University of Rochester Press.
- Butler, J. 2006b/1736. 'The Analogy of Religion, Natural and Revealed, to the Constitution and Course of Nature'. In D. E. White (ed.) *The Works of Bishop Butler*. Rochester: University of Rochester Press, 147–308.
- Bykvist, K. 2002. 'Alternative Actions and the Spirit of Consequentialism'. *Philosophical Studies* 107: 45–68.
- Bykvist, K. 2006. 'Prudence for Changing Selves'. *Utilitas* 18: 264–83.
- Bykvist, K. 2009. 'Objective versus Subjective Moral Oughts'. In L.-G. Johansson, J. Österberg, and R. Śliwiński (eds) *Logic, Ethics and All that Jazz: Essays in Honour of Jordan Howard Sobel*. Uppsala: Uppsala Philosophical Studies, 39–65.
- Byrne, T. 2016. 'Might Anything Be Plain Good?'. *Philosophical Studies* 173: 3335–46.
- Campbell, S. M. 2013. 'An Analysis of Prudential Value'. *Utilitas* 25: 334–54.
- Carlson, E. 1999. 'Consequentialism, Alternatives, and Actualism'. *Philosophical Studies* 96: 253–68.
- Carlyle, T. 1850. *Latter-Day Pamphlets*. London: Chapman and Hall.
- Carson, T. L. 1991. 'A Note on Hooker's "Rule Consequentialism"'. *Mind* 100: 117–21.
- Carson, T. L. 2005. 'Ross and Utilitarianism on Promise Keeping and Lying: Self-Evidence and the Data of Ethics'. *Philosophical Issues* 15: Normativity: 140–57.
- Casal, P. 2007. 'Why Sufficiency Is Not Enough'. *Ethics* 117: 296–326.

- Chang, R., and Sylvan, K. (eds) forthcoming. *The Routledge Handbook of Practical Reasons*. Abingdon: Routledge.
- Chappell, R. Y. 2012. 'Fittingness: The Sole Normative Primitive'. *The Philosophical Quarterly* 62: 684–704.
- Chappell, R. Y. 2015. 'Value Receptacles'. *Noûs* 49: 322–32.
- Chappell, R. Y. 2017. 'Willpower Satisficing'. *Noûs* doi:10.1111/nous.12213.
- Chappell, T. 1993. 'Why God Is Not a Consequentialist'. *Religious Studies* 29: 239–43.
- Chappell, T. (ed.) 2009. *The Problem of Moral Demandingness: New Philosophical Essays*. London: Palgrave Macmillan.
- Charles, D., and Lennon, K. (eds) 1992. *Reduction, Explanation, and Realism*. Oxford: Clarendon Press.
- Chesterton, G. K. 1911. *The Innocence of Father Brown*. London: Cassell and Company.
- Clayton, M., and Williams, A. (eds) 2000. *The Ideal of Equality*. Basingstoke: Macmillan.
- Cocking, D., and Oakley, J. 1995. 'Indirect Consequentialism, Friendship, and the Problem of Alienation'. *Ethics* 106: 86–111.
- Cohen, G. A. 1989. 'On the Currency of Egalitarian Justice'. *Ethics* 99: 906–44.
- Cohen, G. A. 1997. 'Where the Action Is: On the Site of Distributive Justice'. *Philosophy and Public Affairs* 26: 3–30.
- Cohen, J. 2002. 'Deliberation and Democratic Legitimacy'. In D. Estlund (ed.) *Democracy*. Malden: Blackwell, 87–106.
- Colman, A. M., Pulford, B. D., and Rose, J. 2008. 'Collective Rationality in Interactive Decisions: Evidence for Team Reasoning'. *Acta Psychologica* 128: 387–97.
- Conly, S. 1983. 'Utilitarianism and Integrity'. *The Monist* 66: 298–311.
- Cowie, C. 2017. 'Does the Repugnant Conclusion Have Any Probative Force?' *Philosophical Studies* 174: 3021–39.
- Crisp, R. 1992. 'Utilitarianism and the Life of Virtue'. *The Philosophical Quarterly* 42: 139–60.
- Crisp, R. 1998. 'Editor's Introduction', in J. S. Mill, *Utilitarianism*, ed. R. Crisp. Oxford: Oxford University Press.
- Crisp, R. 2000. 'Utilitarianism and Accomplishment'. *Analysis* 60: 264–8.
- Crisp, R. 2001. 'Utilitarianism and Accomplishment Revisited'. *Analysis* 61: 162–4.
- Crisp, R. 2006. *Reasons and the Good*. Oxford: Clarendon Press.
- Crisp, R. (ed.) 2013. *The Oxford Handbook of the History of Ethics*. Oxford: Oxford University Press.
- Crisp, R. (ed.) 2014. *Griffin on Human Rights*. Oxford: Oxford University Press.
- Crisp, R. 2014. 'Taking Stock of Utilitarianism'. *Utilitas* 26: 231–49.
- Crisp, R. 2016. 'Rightness, Parsimony, and Consequentialism: A Response to Peterson'. *Ethical Theory and Moral Practice* 19: 39–47.
- Cruft, R. 2013. 'Why Is It Disrespectful to Violate Rights?' *Proceedings of the Aristotelian Society*, 113: 201–24.
- Cullity, G. 2009. 'Demandingness and Arguments from Presupposition'. In T. Chappell (ed.) *The Problem of Moral Demandingness: New Philosophical Essays*. London: Palgrave Macmillan, 8–34.

- Dancy, J. 2004. *Ethics without Principles*. Oxford: Clarendon Press.
- Daniels N. 1979. 'Wide Reflective Equilibrium and Theory Acceptance in Ethics'. *The Journal of Philosophy* 76: 256–82.
- Darwall, S. (ed.) 1995. *Equal Freedom: Selected Tanner Lectures on Human Values*. Ann Arbor: University of Michigan Press.
- Darwall, S. 2002. *Welfare and Rational Care*. Princeton: Princeton University Press.
- de Lazari-Radek, K. and Singer, P. 2014. *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. Oxford: Oxford University Press.
- Deonna, J. A., and Teroni, F. 2014. 'In What Sense Are Emotions Evaluations?' In C. Todd and S. Roeser (eds) *Emotion and Value*. Oxford: Oxford University Press, 15–31.
- Dickens, C. 2003/1854. *Hard Times*, ed. K. Flint. London: Penguin.
- Diener, E., Diener, M., and Diener, C. 1995. 'Factors Predicting the Subjective Well-Being of Nations'. *Journal of Personality and Social Psychology* 69: 851–64.
- Dietz, A. 2016. 'What We Together Ought to Do'. *Ethics* 126: 955–82.
- Doris, J. M. 1998. 'Persons, Situations, and Virtue Ethics'. *Noûs* 32: 504–30.
- Dorsey, D. 2012. 'Subjectivism without Desire'. *Philosophical Review* 121: 407–42.
- Dorsey, D. 2013. 'Desire-satisfaction and Welfare as Temporal'. *Ethical Theory and Moral Practice* 16: 151–71.
- Dorsey, D. 2017. 'Idealization and the Heart of Subjectivism'. *Noûs* 51: 196–217.
- Dostoevsky, F. 1993/1880. *The Brothers Karamazov*. Translated with an Introduction and Notes by D. McDuff. Harmondsworth: Penguin.
- Doyle, M. W. 1983. 'Kant, Liberal Legacies, and Foreign Affairs'. *Philosophy and Public Affairs* 12: 205–35.
- Driver, J. 1992. 'The Suberogatory'. *Australasian Journal of Philosophy* 70: 286–95.
- Driver, J. 2001. *Uneasy Virtue*. Cambridge: Cambridge University Press.
- Driver, J. 2014. 'Global Utilitarianism'. In B. Eggleston and D. E. Miller (eds) *The Cambridge Companion to Utilitarianism*. Cambridge: Cambridge University Press, 166–76.
- Dworkin, R. 1978. *Taking Rights Seriously*, second impression. London: Duckworth.
- Dworkin, R. 1984. 'Rights as Trumps'. In J. Waldron (ed.) *Theories of Rights*. Oxford: Oxford University Press, 153–67.
- Dworkin, R. 2000. *Sovereign Virtue: The Theory and Practice of Equality*. Cambridge, MA: Harvard University Press.
- Edmundson, W. A. 1998. 'Legitimate Authority without Political Obligation'. *Law and Philosophy* 17: 43–60.
- Eggleston, B., and Miller, D. E. (eds) 2014. *The Cambridge Companion to Utilitarianism*. Cambridge: Cambridge University Press.
- Elster, J. 2013. *Securities against Misrule: Juries, Assemblies, Elections*. Cambridge: Cambridge University Press.
- Endsley, M. R. 1995. 'Towards a Theory of Situation Awareness in Dynamic Systems'. *Human Factors* 37, 32–64.
- Enoch, D. 2005. 'Why Idealize?' *Ethics* 115: 759–87.
- Estlund, D. M. (ed.) 2002. *Democracy*. Malden: Blackwell.

- Estlund, D. M. 2008. *Democratic Authority: A Philosophical Framework*. Princeton: Princeton University Press.
- Ewing, A. C. 1939. 'A Suggested Non-Naturalistic Analysis of Good'. *Mind* 48: 1–22.
- Ewing, A. C. 1948. 'Utilitarianism'. *Ethics* 58: 100–11.
- Feinberg, J. 1970. 'The Nature and Value of Rights'. *The Journal of Value Inquiry* 4: 243–57.
- Feldman, F. 2004. *Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism*. Oxford: Oxford University Press.
- Fischbacher, U., Gächter, S., and Fehr, E. 2001. 'Are People Conditionally Cooperative? Evidence from a Public Goods Experiment'. *Economic Letters* 71: 397–404.
- Fishkin, J. 2014. *Bottlenecks: A New Theory of Equal Opportunity*. New York: Oxford University Press.
- Fletcher, G. 2012. 'The Locative Analysis of Good For Formulated and Defended'. *Journal of Ethics and Social Philosophy* 6: 1–26.
- Fletcher, G. 2013. 'A Fresh Start for the Objective-List Theory of Well-Being'. *Utilitas* 25: 206–20.
- Fletcher, G. 2016. *The Philosophy of Well-Being: An Introduction*. Abingdon: Routledge.
- Fletcher, G. (ed.) 2016. *The Routledge Handbook of Philosophy of Well-Being*. Abingdon: Routledge.
- Fleurbaey, M. 2008. *Fairness, Responsibility, and Welfare*. Oxford: Oxford University Press.
- Foot, P. 1985. 'Utilitarianism and the Virtues'. *Mind* 94: 196–209.
- Forgas, J. P. (ed.) 2000. *Thinking and Feeling: The Role of Affect in Social Cognition*. Cambridge: Cambridge University Press.
- Frankel Paul, E., Miller, F. D., and Paul, J. (eds) 2008. *Utilitarianism: The Aggregation Question*. Cambridge: Cambridge University Press.
- Frankfurt, H. 1987. 'Equality as a Moral Ideal'. *Ethics* 98: 21–43.
- Franklin, D. 2017. 'Calibrating QALYs to Respect Equality of Persons'. *Utilitas* 29: 65–87.
- Frazier, R. L. 1994. 'Act Utilitarianism and Decision Procedures'. *Utilitas* 6: 43–53.
- Freeman, M. (ed.) 2014. *The Future of Children's Rights*. Leiden: Brill Nijhoff.
- Frey, R. G. 1984. 'Act-Utilitarianism, Consequentialism, and Moral Rights'. In R. G. Frey (ed.) *Utility and Rights*. Minneapolis: University of Minnesota Press, 61–85.
- Frey, R. G. (ed.) 1984. *Utility and Rights*. Minneapolis: University of Minnesota Press.
- Galston, W. A. 2010. 'Realism in Political Theory'. *European Journal of Political Theory* 9: 385–411.
- Galvão, P. 2016. 'Rule-Consequentialism and the Significance of Species'. *Utilitas* 28: 396–414.
- Geach, P. T. 1956. 'Good and Evil'. *Analysis* 17: 33–42.
- Gibbard, A. F. 1965. 'Rule-utilitarianism: Merely an Illusory Alternative?' *Australasian Journal of Philosophy* 43: 211–20.
- Gibbard, A. F. 1984. 'Utilitarianism and Human Rights'. *Social Philosophy and Policy* 1: 92–102.

- Gigerenzer, G. 2008. *Rationality for Mortals: How People Cope with Uncertainty*. Oxford: Oxford University Press.
- Gilbert, D. T., and Wilson, T. D. 2000. 'Miswanting: Some Problems in the Forecasting of Future Affective States'. In J. P. Forgas (ed.) *Thinking and Feeling: The Role of Affect in Social Cognition*. Cambridge: Cambridge University Press, 178–97.
- Gilbert, D. T., and Wilson, T. D. 2007. 'Prospection: Experiencing the Future'. *Science* 317: 1351–4.
- Glasgow, J. 2013. 'The Shape of a Life and the Value of Loss and Gain'. *Philosophical Studies* 162: 665–82.
- Goodin, R. E. 1995. *Utilitarianism as a Public Philosophy*. Cambridge: Cambridge University Press.
- Greaves, H. 2016. 'Cluelessness'. *Proceedings of the Aristotelian Society* 116: 311–39.
- Greenawalt, K. 1987. *Conflicts of Law and Morality*. New York: Oxford University Press.
- Griffin, J. 1986. *Well-Being: Its Meaning, Measurement, and Moral Importance*. Oxford: Clarendon Press.
- Griffin, J. 1996. *Value Judgement: Improving our Ethical Beliefs*. Oxford: Clarendon Press.
- Griffin, J. 2008. *On Human Rights*. Oxford: Oxford University Press.
- Hare, R. M. 1952. *The Language of Morals*. Oxford: Clarendon Press.
- Hare, R. M. 1989. *Essays on Political Morality*. Oxford: Clarendon Press.
- Harman, E. 2009. "'I'll Be Glad I Did It" Reasoning and the Significance of Future Desires'. *Philosophical Perspectives* 23: 177–99.
- Harman, G. 2000. *Explaining Value: And Other Essays in Moral Philosophy*. Oxford: Clarendon Press.
- Harsanyi, J. C. 1980. 'Rule Utilitarianism, Rights, Obligations and the Theory of Rational Behavior'. *Theory and Decision* 12: 115–33.
- Hart, H. L. A. 1958. 'Legal and Moral Obligation'. In A. I. Melden (ed.) *Essays in Moral Philosophy*. Seattle, WA: University of Washington Press, 82–107.
- Hart, H. L. A. 1959–60. 'Prolegomenon to the Principles of Punishment'. *Proceedings of the Aristotelian Society*, New Series 60: 1–26.
- Hart, H. L. A. 1961. *The Concept of Law*. Oxford: Clarendon Press.
- Hart, H. L. A. 1982. *Essays on Bentham: Studies in Jurisprudence and Political Theory*. Oxford: Clarendon Press.
- Hausman, D. M. 2012. *Preference, Value, Choice, and Welfare*. New York: Cambridge University Press.
- Hawkins, J. S. 2010. 'The Subjective Intuition'. *Philosophical Studies* 148: 61–8.
- Haybron, D. M. 2008. *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*. Oxford: Oxford University Press.
- Hayward, M. K. 2017. 'Practical Reason, Sympathy and Reactive Attitudes'. *Nous* doi:10.1111/nous.12203.
- Heathwood, C. 2005. 'The Problem of Defective Desires'. *Australasian Journal of Philosophy* 83: 487–504.
- Heathwood, C. 2006. 'Desire Satisfactionism and Hedonism'. *Philosophical Studies* 128: 539–63.

- Heathwood, C. 2016. 'Desire-fulfillment Theory'. In G. Fletcher (ed.) *The Routledge Handbook of Philosophy of Well-Being*. Abingdon: Routledge, 135–47.
- Heuer, U. 2011. 'The Paradox of Deontology, Revisited'. In M. Timmons (ed.) *Oxford Studies in Normative Ethics, Volume 1*. Oxford: Oxford University Press, 236–67.
- Hill, T. 2009. 'Knowing When to Stop'. *American Scientist* 97: 126.
- Hirose, I. 2015. *Moral Aggregation*. New York: Oxford University Press.
- Hobbes, T. 1946/1651. *Leviathan*, ed. M. Oakshott. Oxford: Basil Blackwell.
- Hohfeld, W. N. 1913. 'Some Fundamental Legal Conceptions as Applied in Judicial Reasoning'. *The Yale Law Journal* 23: 16–59.
- Holroyd, J. 2007. 'A Communicative Conception of Moral Appraisal'. *Ethical Theory and Moral Practice* 10: 267–78.
- Hooker, B. 1991. 'Theories of Welfare, Theories of Good Reasons for Action, and Ontological Naturalism'. *Philosophical Papers* 20: 25–36.
- Hooker, B. (ed.) 1993. *Rationality, Rules, and Utility: New Essays on the Moral Philosophy of Richard B. Brandt*. Boulder, CO: Westview Press.
- Hooker, B. 2000. *Ideal Code, Real World. A Rule-consequentialist Theory of Morality*. Oxford: Clarendon Press.
- Hooker, B. 2009. 'The Demandingness Objection'. In T. Chappell (ed.) *The Problem of Moral Demandingness: New Philosophical Essays*. London: Palgrave Macmillan, 148–62.
- Hooker, B. 2014. 'Griffin on Human Rights'. In R. Crisp (ed.) *Griffin on Human Rights*. Oxford: Oxford University Press, 170–84.
- Hooker, B. 2015. 'The Elements of Well-Being'. *Journal of Practical Ethics* 3: 15–35.
- Hooker, B., Mason, E., and Miller, D. (eds) 2000. *Morality, Rules, and Consequences: A Critical Reader*. Edinburgh: Edinburgh University Press.
- Howard, C. forthcoming. 'The Fundamentality of Fit'. In R. Shafer-Landau (ed.) *Oxford Studies in Metaethics, Volume 14*. Oxford: Oxford University Press.
- Howard-Snyder, F. 1993. 'Rule Consequentialism Is a Rubber Duck'. *American Philosophical Quarterly* 30: 271–8.
- Howard-Snyder, F. 1997. 'The Rejection of Objective Consequentialism'. *Utilitas* 9: 241–8.
- Hsieh, C. C., and Pugh, M. D. 1993. 'Poverty, Income Inequality, and Violent Crime: A Meta-analysis of Recent Aggregate Data Studies'. *Criminal Justice Review* 18: 182–202.
- Huemer, M. 2008. 'In Defence of Repugnance'. *Mind* 117: 899–933.
- Huppert, F. A., Baylis, N., and Keverne, B. (eds) 2005. *The Science of Well-Being*. Oxford: Oxford University Press.
- Hurka, T. 1993. *Perfectionism*. New York: Oxford University Press.
- Hurka, T. 2001. *Virtue, Vice, and Value*. New York: Oxford University Press.
- Hurka, T. 2014. *British Ethical Theorists from Sidgwick to Ewing*. Oxford: Oxford University Press.
- Hurley, S. L. 1989. *Natural Reasons: Personality and Polity*. New York: Oxford University Press.
- Hurley, S. L. 2001. 'Reason and Motivation: The Wrong Distinction?' *Analysis* 61: 151–5.

- Jackson, F. 1987. 'Group Morality'. In P. Pettit, R. Sylvan, and J. Norman (eds) *Metaphysics and Morality: Essays in Honour of J. C. C. Smart*. Oxford: Basil Blackwell, 91–110.
- Jackson, F. 1991. 'Decision-theoretic Consequentialism and the Nearest and Dearest Objection'. *Ethics* 101: 461–82.
- Jackson, F., and Pargetter, R. 1986. 'Oughts, Options, and Actualism'. *Philosophical Review* 95: 233–55.
- Jamieson, D. 2014. *Reason in a Dark Time: Why the Struggle against Climate Change Failed—and What It Means for our Future*. New York: Oxford University Press.
- Jeske, D. 1998. 'Families, Friends, and Special Obligations'. *Canadian Journal of Philosophy* 28: 527–55.
- Johansson, L.-G., Österberg, J., and Śliwiński, R. (eds) 2009. *Logic, Ethics and All that Jazz: Essays in Honour of Jordan Howard Sobel*. Uppsala: Uppsala Philosophical Studies.
- Kagan, S. 1989. *The Limits of Morality*. Oxford: Clarendon Press.
- Kagan, S. 1998. *Normative Ethics*. Boulder: Westview.
- Kagan, S. 2000. 'Evaluative Focal Points'. In B. Hooker, E. Mason, and D. Miller (eds) *Morality, Rules and Consequences: A Critical Reader*. Edinburgh: Edinburgh University Press, 134–55.
- Kagan, S. 2009. 'Well-Being as Enjoying the Good'. *Philosophical Perspectives* 23: 253–72.
- Kagan, S. 2012. *The Geometry of Desert*. New York: Oxford University Press.
- Kagan, S. 2014. 'An Introduction to Ill-Being'. In M. Timmons (ed.) *Oxford Studies in Normative Ethics, Volume 4*. Oxford: Oxford University Press, 261–88.
- Kahneman, D. 2011. *Thinking, Fast and Slow*. London: Penguin.
- Kahneman, D., Diener, E., and Schwartz, N. (eds) 1999. *Well-Being: The Foundations of Hedonic Psychology*. New York: Russell Sage Foundation.
- Kamm, F. 2007. *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*. New York: Oxford University Press.
- Kant, I. 1996a. *Practical Philosophy*, ed. and trans. M. Gregor. Cambridge: Cambridge University Press.
- Kant, I. 1996b/1797. 'On a Supposed Right to Lie from Philanthropy'. In M. Gregor (ed. and trans.) *Practical Philosophy*. Cambridge: Cambridge University Press., 611–15.
- Keller, S. 2009. 'Welfarism'. *Philosophy Compass* 4: 82–95.
- Kittay, E. F. 2005. 'At the Margins of Moral Personhood'. *Ethics* 116: 100–31.
- Kolodny, N., and Wallace, R. J. 2003. 'Promises and Practices Revisited'. *Philosophy and Public Affairs* 31: 119–54.
- Korsgaard, C. M. 1996. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.
- Kraut, R. 1994. 'Desire and the Human Good'. *Proceedings and Addresses of the American Philosophical Association* 68: 39–54.
- Kraut, R. 2007. *What Is Good and Why: The Ethics of Well-Being*. Cambridge, MA: Harvard University Press.
- Kubala, R. 2017. 'Valuing and Believing Valuable'. *Analysis* 77: 59–65.

- Kymlicka, W. 2002. *Contemporary Political Philosophy: An Introduction*, second ed. Oxford: Oxford University Press.
- Lang, G. 2013. 'Should Utilitarianism Be Scalar?' *Utilitas* 25: 80–95.
- Law, I. 1999. 'Rule-Consequentialism's Dilemma'. *Ethical Theory and Moral Practice* 2: 263–75.
- Lawlor, R. 2009. 'The Rejection of Scalar Consequentialism'. *Utilitas* 21: 100–16.
- Leavis, F. R. 1950. 'Introduction'. In J. S. Mill, *On Bentham and Coleridge*. London: Chatto and Windus, 1–38.
- Lenman, J. 2000. 'Consequentialism and Cluelessness'. *Philosophy and Public Affairs* 29: 342–70.
- Lerner, A. P. 1944. *The Economics of Control: Principles of Welfare Economics*. New York: Macmillan.
- Lerner, A. P. 1978. 'Utilitarian Marginalism (Nozick, Rawls, Justice, and Welfare)'. *Eastern Economic Journal* 4: 51–65.
- Lewis, D. 1989. 'Dispositional Theories of Value'. *Proceedings of the Aristotelian Society*, Supplementary Volume 63: 113–37.
- Lillehammer, H. 2007. 'Davidson on Value and Objectivity'. *Dialectica* 61: 203–17.
- Lin, E. 2017a. 'Against Welfare Subjectivism'. *Noûs* 51: 354–77.
- Lin, E. 2017b. 'Enumeration and Explanation in Theories of Welfare'. *Analysis* 77: 65–73.
- Lively, J. and Rees, J. (eds) 1978. *Utilitarian Logic and Politics: James Mill's 'Essay on Government', Macaulay's Critique, and the Ensuing Debate*. Oxford: Clarendon Press.
- Locke, J. 1988/1690. *Two Treatises of Government*, ed. P. Laslett. Cambridge: Cambridge University Press.
- Loewenstein, G., O'Donoghue, T., and Rabin, M. 2003. 'Projection Bias in Predicting Future Utility'. *The Quarterly Journal of Economics* 118: 1209–48.
- Lorant, V., Delière, D., Eaton, W., Robert, A., Philippot, P., and Ansseau, M. 2003. 'Socioeconomic Inequalities in Depression: A Meta-analysis'. *American Journal of Epidemiology* 157: 98–112.
- Lord, E. 2015. 'Acting for the Right Reasons, Abilities, and Obligation'. In R. Shafer-Landau (ed.) *Oxford Studies in Metaethics, Volume 14*. Oxford: Oxford University Press, 26–52.
- Louise, J. 2004. 'Relativity of Value and the Consequentialist Umbrella'. *The Philosophical Quarterly* 54: 518–36.
- Lucas, R. E. 2007. 'Adaptation and the Set-Point Model of Subjective Well-Being: Does Happiness Change after Major Life Events?' *Current Directions in Psychological Science* 16: 75–9.
- Lynch, J. W., Kaplan, G. A., Pamuk, E. R., Cohen, R. D., Heck, K. E., Balfour, J. L., and Yen, I. H. 1998. 'Income Inequality and Mortality in Metropolitan Areas of the United States'. *American Journal of Public Health* 88: 1074–80.
- Lyons, D. 1965. *Forms and Limits of Utilitarianism*. Oxford: Clarendon Press.
- Lyons, D. 1994. *Rights, Welfare, and Mill's Moral Theory*. New York: Oxford University Press.
- MacAskill, W. 2014. 'Replaceability, Career Choice, and Making a Difference'. *Ethical Theory and Moral Practice* 17: 269–83.
- Mackie, G. 2003. *Democracy Defended*. Cambridge: Cambridge University Press.

- Martin, W. L., Murray, P. S., Bates, P. R., and Lee, P. S. Y. 2015. 'Fear-Potentiated Startle: A Review from an Aviation Perspective'. *The International Journal of Aviation Psychology*, 25: 97–107.
- Mason, E. 1998. 'Can an Indirect Consequentialist Be a Real Friend?' *Ethics* 108: 386–93.
- Mason, E. 2004. 'Consequentialism and the Principle of Indifference'. *Utilitas* 16: 316–21.
- Mason, E. 2013. 'Objectivism and Prospectivism about Rightness'. *Journal of Ethics and Social Philosophy* 7: 1–21.
- Matravers, M. 2000. *Justice and Punishment: The Rationale of Coercion*. Oxford: Oxford University Press.
- Maynard, J. L., and Worsnip, A. 2018. 'Is There a Distinctively Political Normativity?' *Ethics* 128: 756–87.
- McClennen, E. F. 1990. *Rationality and Dynamic Choice: Foundational Explorations*. Cambridge: Cambridge University Press.
- McCloskey, H. J. 1967. 'Utilitarian and Retributive Punishment'. *The Journal of Philosophy* 64: 91–110.
- McElwee, B. 2010a. 'Consequentialism and Permissibility'. *Utilitas* 22: 171–83.
- McElwee, B. 2010b. 'The Rights and Wrongs of Consequentialism'. *Philosophical Studies* 151: 393–412.
- McElwee, B. 2015. 'The Value of the Virtues'. *Utilitas* 27: 61–81.
- McHugh, C., and Way, J. 2016. 'Fittingness First'. *Ethics* 126: 575–606.
- McKerlie, D. 1996. 'Equality'. *Ethics* 106: 274–96.
- McKerlie, D. 2013. *Justice between the Young and the Old*. Oxford: Oxford University Press.
- McLean, I. 1987. *Public Choice: An Introduction*. Oxford: Basil Blackwell.
- McMahan, J. 1996. 'Cognitive Disability, Misfortune, and Justice'. *Philosophy and Public Affairs* 25: 3–35.
- McMahon, C. 1991. 'The Paradox of Deontology'. *Philosophy and Public Affairs* 20: 350–77.
- McNaughton, D., and Rawling, P. 2001. 'Achievement, Welfare and Consequentialism'. *Analysis* 61: 156–62.
- Melden, A. I. (ed.) 1958. *Essays in Moral Philosophy*. Seattle, WA: University of Washington Press.
- Mill, J. 1978/1820. 'Essay on Government'. In J. Lively and J. Rees (eds) *Utilitarian Logic and Politics: James Mill's 'Essay on Government', Macaulay's Critique, and the Ensuing Debate*. Oxford: Clarendon Press, 53–95.
- Mill, J. S. 1950. *On Bentham and Coleridge*. Intr. F. R. Leavis. London: Chatto and Windus.
- Mill, J. S. 1991a. *On Liberty and Other Essays*, ed. J. Gray. Oxford: Oxford University Press.
- Mill, J. S. 1991b/1859. 'On Liberty'. In J. Gray (ed.) *On Liberty and Other Essays*. Oxford: Oxford University Press, 1–128.
- Mill, J. S. 1991c/1861. 'Considerations on Representative Government'. In J. Gray (ed.) *On Liberty and Other Essays*. Oxford: Oxford University Press, 203–467.
- Mill, J. S. 1998/1861. *Utilitarianism*, ed. R. Crisp. Oxford: Oxford University Press.

- Mill, J. S. 2008/1848. *Principles of Political Economy: Books III–V and Appendices*, ed. J. M. Robson. London: Routledge.
- Miller, D. 1992. 'Deliberative Democracy and Social Choice'. *Political Studies* 40: 54–67. Reprinted in D. Estlund (ed.) *Democracy*. Malden: Blackwell, 289–307.
- Miller, D. 2016. 'In What Sense Must Political Philosophy Be Political?' *Social Philosophy and Policy* 33: 155–74.
- Miller, D. E. 2003. 'Actual-Consequence Act Utilitarianism and the Best Possible Humans'. *Ratio (new series)* 16: 49–62.
- Moore, A., and Crisp, R. 1996. 'Welfarism in Moral Theory'. *Australasian Journal of Philosophy* 74: 598–613.
- Moore, G. E. 1993/1903. *Principia Ethica*, rev. ed., ed. T. Baldwin. Cambridge: Cambridge University Press.
- Moore, G. E. 2005. *Ethics, with 'The Nature of Moral Philosophy'*, ed. W. H. Shaw. Oxford: Clarendon Press.
- Mulgan, T. 2000. 'Two Moral Counterfactuals'. *Philosophical Forum* 31: 47–55.
- Mulgan, T. 2001. *The Demands of Consequentialism*. Oxford: Clarendon Press.
- Mulgan, T. 2006. *Future People: A Moderate Consequentialist Account of our Obligations to Future Generations*. Oxford: Clarendon Press.
- Mulgan, T. 2011. *Ethics for a Broken World: Imagining Philosophy after Catastrophe*. Durham: Acumen.
- Mulgan, T. 2015. *Purpose in the Universe: The Moral and Metaphysical Case for Ananthropocentric Purposivism*. Oxford: Oxford University Press.
- Mulgan, T. 2017. 'How Should Utilitarians Think about the Future?' *Canadian Journal of Philosophy* 47: 290–312.
- Murphy, L. B. 1999. 'Institutions and the Demands of Justice'. *Philosophy and Public Affairs* 27: 251–91.
- Murphy, L. B., and Nagel, T. 2002. *The Myth of Ownership: Taxes and Justice*. New York: Oxford University Press.
- Murphy, M. C. 1999. 'The Simple Desire-Fulfillment Theory'. *Noûs* 33: 247–72.
- Myrdal, G. 1953. *The Political Element in the Development of Economic Theory*, trans. P. Streeten. Abingdon: Routledge.
- Nelson, M. T. 1991. 'Utilitarian Eschatology'. *American Philosophical Quarterly* 28: 339–47.
- Nelson, M. T. 2015. 'What the Utilitarian Cannot Think'. *Ethical Theory and Moral Practice* 18: 717–29.
- Nietzsche, F. 2002/1886. *Beyond Good and Evil*, ed. R.-P. Horstmann and J. Norman, trans. J. Norman. Cambridge: Cambridge University Press.
- Norcross, A. 2006. 'The Scalar Approach to Utilitarianism'. In H. R. West (ed.) *The Blackwell Guide to Mill's Utilitarianism*. Oxford: Oxford University Press, 217–32.
- Nozick, R. 1974. *Anarchy, State, and Utopia*. Oxford: Blackwell.
- Oberdiek, J. 2004. 'Lost in Moral Space: On the Infringing/Violating Distinction and Its Place in the Theory of Rights'. *Law and Philosophy* 23: 325–46.
- O'Hear, A. (ed.) 2006. *Political Philosophy. Royal Institute of Philosophy Supplement*. Cambridge: Cambridge University Press.
- Okin, S. M. 1989. *Justice, Gender, and the Family*. New York: Basic Books.

- O'Neill, M. 2008. 'What Should Egalitarians Believe?' *Philosophy and Public Affairs* 36: 119–56.
- O'Neill, O. 2005. 'The Dark Side of Human Rights'. *International Affairs* 81: 427–39.
- Orr, S. W. 2007. 'Values, Preferences, and the Citizen–Consumer Distinction in Cost–Benefit Analysis'. *Politics, Philosophy and Economics* 6: 107–30.
- Orsi, F. 2015. 'The Guise of the Good'. *Philosophy Compass* 10: 714–24.
- Otsuka, M., and Voorhoeve, A. 2009. 'Why It Matters That Some Are Worse Off Than Others: An Argument against the Priority View'. *Philosophy and Public Affairs* 37: 171–99.
- Ott, J. 2005. 'Level and Inequality of Happiness in Nations: Does Greater Happiness of a Greater Number Imply Greater Inequality of Happiness?' *Journal of Happiness Studies* 6: 397–420.
- Owens, D. 2012. *Shaping the Normative Landscape*. Oxford: Oxford University Press.
- Page, S. E. 2011. *Diversity and Complexity*. Princeton, NJ: Princeton University Press.
- Parfit, D. 1987. *Reasons and Persons*, reprinted with corrections. Oxford: Clarendon Press.
- Parfit, D. 2000. 'Equality or Priority?' In M. Clayton and A. Williams (eds) *The Ideal of Equality*. Basingstoke: Macmillan, 81–125.
- Parfit, D. 2011. *On What Matters*, vol. 1, ed. S. Scheffler. Oxford: Oxford University Press.
- Paul, L. A. 2014. *Transformative Experience*. Oxford: Oxford University Press.
- Pedersen, J. T. 1982. 'On the Educational Function of Political Participation: A Comparative Analysis of John Stuart Mill's Theory and Contemporary Survey Research Findings'. *Political Studies* 30: 557–68.
- Peter, F. 2007. 'Democratic Legitimacy and Proceduralist Social Epistemology'. *Politics, Philosophy and Economics* 6: 329–53.
- Peterson, M. 2013. *The Dimensions of Consequentialism: Ethics, Equality and Risk*. Cambridge: Cambridge University Press.
- Pettigrew, R. 2015. 'Transformative Experience and Decision Theory'. *Philosophy and Phenomenological Research* 91: 766–74.
- Pettit, P. 1988. 'The Consequentialist Can Recognise Rights'. *The Philosophical Quarterly* 38: 42–55.
- Pettit, P. 1996. 'Freedom as Antipower'. *Ethics* 106: 576–604.
- Pettit, P. 1997. *Republicanism: A Theory of Freedom and Government*. Oxford: Clarendon Press.
- Pettit, P. 2001. 'Deliberative Democracy and the Discursive Dilemma'. *Philosophical Issues* 11: 268–99.
- Pettit, P., and Smith, M. 2000. 'Global Consequentialism'. In B. Hooker, E. Mason, and D. Miller (eds) *Morality, Rules, and Consequences: A Critical Reader*. Edinburgh: Edinburgh University Press, 121–33.
- Pettit, P., Sylvan, R., and Norman, J. (eds) 1987. *Metaphysics and Morality. Essays in Honour of J. J. C. Smart*. Oxford: Basil Blackwell.
- Podgorski, A. 2018. 'Wouldn't It Be Nice? Moral Rules and Distant Worlds'. *Noûs* 52: 279–94.

- Pogge, T. W. 1995. 'Utilitarianism and Equality'. *Midwest Studies in Philosophy* XX: 332–9.
- Portmore, D. W. 2011. *Commonsense Consequentialism: Wherein Morality Meets Rationality*. New York: Oxford University Press.
- Portmore, D. W. forthcoming. *Opting for the Best: Oughts and Options*.
- Posner, R. A. 1979. 'Utilitarianism, Economics, and Legal Theory'. *The Journal of Legal Studies* 8: 103–40.
- Postema, G. J. 1998. 'Bentham's Equality-Sensitive Utilitarianism'. *Utilitas* 10: 144–58.
- Railton, P. 1984. 'Alienation, Consequentialism, and the Demands of Morality'. *Philosophy and Public Affairs* 13: 134–71.
- Railton, P. 2003a. *Facts, Values, and Norms: Essays toward a Morality of Consequence*. Cambridge: Cambridge University Press.
- Railton, P. 2003b. 'Facts and Values'. In P. Railton, *Facts, Values, and Norms: Essays toward a Morality of Consequence*. Cambridge: Cambridge University Press, 43–68.
- Railton, P. 2012. 'That Obscure Object, Desire'. *Proceedings and Addresses of the American Philosophical Association* 86: 22–46.
- Rawls, J. 1955. 'Two Concepts of Rules'. *The Philosophical Review* 64: 3–32.
- Rawls, J. 1996. *Political Liberalism*. New York: Columbia University Press.
- Rawls, J. 1997. 'The Idea of Public Reason Revisited'. *The University of Chicago Law Review* 64: 765–807.
- Rawls, J. 1999a. *A Theory of Justice*, revised edition. Oxford: Oxford University Press.
- Rawls, J. 1999b. *The Law of Peoples, with 'The Idea of Public Reason Revisited'*. Cambridge, MA: Harvard University Press.
- Rawls, J. 1999c. *Collected Papers*, ed. S. Freeman. Cambridge, MA: Harvard University Press.
- Rawls, J. 1999d/1958. 'Justice as Fairness'. In S. Freeman (ed.) *Collected Papers*. Cambridge, MA: Harvard University Press, 47–72.
- Raz, J. 1984. 'On the Nature of Rights'. *Mind* 93: 194–214.
- Raz, J. 1986. *The Morality of Freedom*. Oxford: Clarendon Press.
- Regan, D. 1980. *Utilitarianism and Co-operation*. Oxford: Clarendon Press.
- Rice, C. M. 2013. 'Defending the Objective List Theory of Well-Being'. *Ratio (new series)* 26: 196–211.
- Ridge, M. 2006. 'Introducing Variable Rate Rule Utilitarianism'. *The Philosophical Quarterly* 56: 242–53.
- Riley, J. 1990. 'Utilitarian Ethics and Democratic Government'. *Ethics* 100: 335–48.
- Roemer, J. E. 1996. *Theories of Distributive Justice*. Cambridge, MA: Harvard University Press.
- Rosati, C. S. 1995. 'Persons, Perspectives, and Full Information Accounts of the Good'. *Ethics* 105: 296–325.
- Rosati, C. S. 1996. 'Internalism and the Good for a Person'. *Ethics* 106: 297–326.
- Rosato, S. 2003. 'The Flawed Logic of Democratic Peace Theory'. *The American Political Science Review* 97: 585–602.
- Ross, W. D. 2002/1930. *The Right and the Good*, ed. P. Stratton-Lake. Oxford: Clarendon Press.

- Rossi, E. 2012. 'Justice, Legitimacy and (Normative) Authority for Political Realists'. *Critical Review of International Social and Political Philosophy* 15: 149–64.
- Runciman, D. 2013. *The Confidence Trap: A History of Democracy in Crisis from World War I to the Present*. Princeton, NJ: Princeton University Press.
- Sarch, A. F. 2011. 'Internalism about a Person's Good: Don't Believe It'. *Philosophical Studies* 154: 161–84.
- Sarch, A. F. 2012. 'Multi-Component Theories of Well-Being and their Structure'. *Pacific Philosophical Quarterly* 93: 439–71.
- Sartorius, R. E. 1975. *Individual Conduct and Social Norms: A Utilitarian Account of Social Union and the Rule of Law*. Encino: Dickenson.
- Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, MA: Belknap Press.
- Scheffler, S. 1988. 'Introduction', in S. Scheffler (ed.) *Consequentialism and its Critics*. Oxford: Oxford University Press, 1–13.
- Scheffler, S. (ed.) 1988. *Consequentialism and its Critics*. Oxford: Oxford University Press.
- Scheffler, S. 1992. *Human Morality*. New York: Oxford University Press.
- Scheffler, S. 1994. *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*, rev. ed. Oxford: Clarendon Press.
- Schofield, P. 2006. *Utility and Democracy: The Political Thought of Jeremy Bentham*. Oxford: Oxford University Press.
- Schroeder, M. 2007a. *Slaves of the Passions*. Oxford: Oxford University Press.
- Schroeder, M. 2007b. 'Teleology, Agent-Relative Value, and "Good"'. *Ethics* 117: 265–95.
- Schroeder, S. A. 2017. 'Consequentializing and its Consequences'. *Philosophical Studies* 174: 1475–97.
- Schultz, B. 2017. *The Happiness Philosophers: The Lives and Works of the Great Utilitarians*. Princeton, NJ: Princeton University Press.
- Sen, A. K. 1979. 'Utilitarianism and Welfarism'. *The Journal of Philosophy* 76: 463–89.
- Sen, A. K. 1981. *Poverty and Famines: An Essay on Entitlement and Deprivation*. Oxford: Clarendon Press.
- Sen, A. K. 1982. 'Rights and Agency'. *Philosophy and Public Affairs* 11: 3–39.
- Sen, A. K. 1995. 'Equality of What?' In S. Darwall (ed.) *Equal Freedom: Selected Tanner Lectures on Human Values*. Ann Arbor, MI: University of Michigan Press, 307–30.
- Sen, A. K. 1999. *Development as Freedom*. Oxford: Oxford University Press.
- Sen, A. K., and Williams, B. A. O. (eds) 1982a. *Utilitarianism and beyond*. Cambridge: Cambridge University Press.
- Sen, A. K., and Williams, B. A. O. 1982b. 'Introduction: Utilitarianism and beyond', in A. Sen and B. A. O. Williams (eds) *Utilitarianism and beyond*. Cambridge: Cambridge University Press, 1–21.
- Shafer-Landau, R. (ed.) 2015. *Oxford Studies in Metaethics, Volume 10*. Oxford: Oxford University Press.
- Shafer-Landau, R. (ed.) forthcoming. *Oxford Studies in Metaethics, Volume 14*. Oxford: Oxford University Press.

- Shaver, R. 2004. 'The Appeal of Utilitarianism'. *Utilitas* 16: 235–50.
- Shaw, W. H. 1993. 'Welfare, Equality, and Distribution: Brandt from the Left'. In B. Hooker (ed.) *Rationality, Rules, and Utility: New Essays on the Moral Philosophy of Richard B. Brandt*. Boulder, CO: Westview Press, 165–87.
- Shaw, W. H. 1999. *Contemporary Ethics: Taking Account of Utilitarianism*. Malden: Blackwell.
- Shields, L. 2012. 'The Prospects for Sufficiencyarianism'. *Utilitas* 24: 101–17.
- Sidgwick, H. 1907. *The Methods of Ethics*, seventh edition. London: Macmillan.
- Simmons, A. J. 1979. *Moral Principles and Political Obligations*. Princeton, NJ: Princeton University Press.
- Singer, P. 2005. 'Ethics and Intuitions'. *The Journal of Ethics* 9: 331–52.
- Sinnott-Armstrong, W. 1984. "'Ought" Conversationally Implies "Can"'. *Philosophical Review* 93: 249–61.
- Smart, J. J. C. 1973. 'An Outline of a System of Utilitarian Ethics'. In J. J. C. Smart and B. A. O. Williams, *Utilitarianism: for and against*. Cambridge: Cambridge University Press, 1–74.
- Smart, J. J. C., and Williams, B. A. O. 1973. *Utilitarianism: for and against*. Cambridge: Cambridge University Press.
- Smith, D. A., and Tolbert, C. 2007. 'The Instrumental and Educative Effects of Ballot Measures: Research on Direct Democracy in the American States'. *State Politics and Policy Quarterly* 7: 416–45.
- Smith, H. M. 2010a. 'Measuring the Consequences of Rules'. *Utilitas* 22: 413–33.
- Smith, H. M. 2010b. 'Subjective Rightness'. *Social Philosophy and Policy* 27: 64–110.
- Smith, M. A. 1992. 'Valuing: Desiring or Believing?' In D. Charles and K. Lennon (eds) *Reduction, Explanation, and Realism*. Oxford: Clarendon Press, 323–59.
- Smith, M. A. 1994. *The Moral Problem*. Oxford: Blackwell.
- Snow, N. (ed.) 2018. *The Oxford Handbook of Virtue*. Oxford: Oxford University Press.
- Sobel, D. 2007. 'The Impotence of the Demandingness Objection'. *Philosopher's Imprint* 7: 1–17.
- Sobel, D. 2009. 'Subjectivism and Idealization'. *Ethics* 119: 336–52.
- Sobel, D. 2016. *From Valuing to Value: Towards a Defense of Subjectivism*. Oxford: Oxford University Press.
- Sreenivasan, G. 2002. 'Errors about Errors: Virtue Theory and Trait Attribution'. *Mind* 111: 47–68.
- Stein, M. S. 2006. *Distributive Justice and Disability: Utilitarianism against Egalitarianism*. New Haven, CT: Yale University Press.
- Steiner, H. 1994. *An Essay on Rights*. Oxford: Blackwell.
- Steinhoff, U. (ed.) 2015. *Do All Persons Have Equal Moral Worth? On 'Basic Equality' and Equal Respect and Concern*. Oxford: Oxford University Press.
- Streumer, B. 2003. 'Does "Ought" Conversationally Implicate "Can"?' *European Journal of Philosophy* 11: 219–28.
- Streumer, B. 2007. 'Reasons and Impossibility'. *Philosophical Studies* 136: 351–84.
- Su, H. 2013. *Economic Justice and Liberty: The Social Philosophy in John Stuart Mill's Utilitarianism*. Abingdon: Routledge.

- Sugden, R. 1993. 'Thinking as a Team: Towards an Explanation of Nonselish Behavior'. *Social Philosophy and Policy* 10: 69–89.
- Sumner, L. W. 1996. *Welfare, Happiness, and Ethics*. Oxford: Clarendon Press.
- Swanton, C. 2003. *Virtue Ethics: A Pluralistic View*. Oxford: Oxford University Press.
- Tasioulas, J. 2013. 'Justice, Equality, and Rights'. In R. Crisp (ed.) *The Oxford Handbook of the History of Ethics*. Oxford: Oxford University Press, 768–92.
- Temkin, L. S. 2000. 'Equality, Priority, and the Levelling Down Objection'. In M. Clayton and A. Williams (eds) *The Ideal of Equality*. Basingstoke: Macmillan, 126–61.
- Temkin, L. S. 2012. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. New York: Oxford University Press.
- Tenenbaum, S. forthcoming. 'The Guise of the Good'. In R. Chang and K. Sylvan (eds) *The Routledge Handbook of Practical Reasons*. Abingdon: Routledge.
- Tetlock, P., and Gardner, D. 2016. *Superforecasting: The Art and Science of Prediction*. London: Random House.
- Thomson, J. J. 1990. *The Realm of Rights*. Cambridge, MA: Harvard University Press.
- Thomson, J. J. 1997. 'The Right and the Good'. *The Journal of Philosophy* 94: 273–98.
- Tiberius, V. 2004. 'Cultural Differences and Philosophical Accounts of Well-Being'. *Journal of Happiness Studies* 5: 293–314.
- Tiberius, V. 2008. *The Reflective Life: Living Wisely with our Limits*. Oxford: Oxford University Press.
- Timmons, M. (ed.) 2011. *Oxford Studies in Normative Ethics, Volume 1*. Oxford: Oxford University Press.
- Timmons, M. (ed.) 2014. *Oxford Studies in Normative Ethics, Volume 4*. Oxford: Oxford University Press.
- Timmons, M. (ed.) 2016. *Oxford Studies in Normative Ethics, Volume 6*. Oxford: Oxford University Press.
- Tobia, K. 2013. 'Rule Consequentialism and the Problem of Partial Acceptance'. *Ethical Theory and Moral Practice* 16: 643–52.
- Tobia, K. 2017. 'A Defense of Scalar Utilitarianism'. *American Philosophical Quarterly* 54: 283–94.
- Todd, C. and Roeser, S. (eds) 2014. *Emotion and Value*. Oxford: Oxford University Press.
- Toppinen, T. 2016. 'Rule Consequentialism (and Kantian Contractualism) at Top Rates'. *The Philosophical Quarterly* 66: 122–35.
- Turri, J. 2010. 'On the Relationship between Propositional and Doxastic Justification'. *Philosophy and Phenomenological Research* 80: 312–26.
- Tyler, T. R. 1990. *Why People Obey the Law*. New Haven, CT: Yale University Press.
- Tyler, T. R., and Darley, J. M. 2000. 'Building a Law-Abiding Society: Taking Public Views about Morality and the Legitimacy of Legal Authorities into Account When Formulating Substantive Law'. *Hofstra Law Review* 28: 707–39.
- Tyler, T. R., and Mitchell, G. 1994. 'Legitimacy and the Empowerment of Discretionary Legal Authority: The United States Supreme Court and Abortion Rights'. *Duke Law Journal* 43: 703–814.

- Ullmann-Margalit, E. 2006. 'Big Decisions: Opting, Converting, Drifting'. *Royal Institute of Philosophy Supplement* 58: 157–72.
- van Ackeren, M., and Sticker, M. 2015. 'Kant and Moral Demandingness'. *Ethical Theory and Moral Practice* 18: 75–89.
- Velleman, D. 1991. 'Well-Being and Time'. *Pacific Philosophical Quarterly* 72: 48–77.
- Velleman, D. 1992. 'The Guise of the Good'. *Noûs* 26: 3–26.
- Vessel, J.-P. 2010. 'Supererogation for Utilitarianism'. *American Philosophical Quarterly* 47: 299–319.
- Viroli, M. 2014. *Redeeming The Prince: The Meaning of Machiavelli's Masterpiece*. Princeton, NJ: Princeton University Press.
- Višák, T. 2013. *Killing Happy Animals: Explorations in Utilitarian Ethics*. London: Palgrave Macmillan.
- Višák, T., and Garner, R. (eds) 2016. *The Ethics of Killing Animals*. Oxford: Oxford University Press.
- Waldron, J. (ed.) 1984. *Theories of Rights*. Oxford: Oxford University Press.
- Walzer, M. 1973. 'Political Action: The Problem of Dirty Hands'. *Philosophy and Public Affairs* 2: 160–80.
- Watson, G. 1975. 'Free Agency'. *The Journal of Philosophy* 72: 205–20.
- Webber, J. 2006. 'Virtue, Character and Situation'. *Journal of Moral Philosophy* 3: 193–213.
- Weber, M. 1994a. *Weber: Political Writings*, ed. P. Lassman and R. Speirs. Cambridge: Cambridge University Press.
- Weber, M. 1994b/1919. 'The Profession and Vocation of Politics'. In P. Lassman and R. Speirs (eds) *Weber: Political Writings*, Cambridge: Cambridge University Press, 309–69.
- Wellman, C. H. 2001. 'Toward a Liberal Theory of Political Obligation'. *Ethics* 111: 735–59.
- Wenar, L. 2005. 'The Nature of Rights'. *Philosophy and Public Affairs* 33: 223–52.
- Wertheimer, A. 1987. *Coercion*. Princeton, NJ: Princeton University Press.
- West, H. R. (ed.) 2006. *The Blackwell Guide to Mill's Utilitarianism*. Oxford: Blackwell.
- Willand, E. 2005. 'Monkeys, Typewriters, and Objective Consequentialism'. *Ratio (new series)* 18: 352–60.
- Williams, B. A. O. 1973. 'A Critique of Utilitarianism'. In J. J. C. Smart and B. A. O. Williams, *Utilitarianism: for and against*. Cambridge: Cambridge University Press, 75–150.
- Williams, B. A. O. 1981a. *Moral Luck: Philosophical Papers 1973–1980*. Cambridge: Cambridge University Press.
- Williams, B. A. O. 1981b. 'Internal and External Reasons'. In B. A. O. Williams *Moral Luck: Philosophical Papers 1973–1980*. Cambridge: Cambridge University Press, 101–13.
- Williams, B. A. O. 1985. *Ethics and the Limits of Philosophy*. London: Fontana.
- Williams, B. A. O. 2005. *In the Beginning Was the Deed: Realism and Moralism in Political Argument*, ed. G. Hawthorn. Princeton, NJ: Princeton University Press.

- Wilson, T. D. 2002. *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, MA: Harvard University Press.
- Wilson, T. D., Lindsey, S., and Schooler, T. Y. 2000. 'A Model of Dual Attitudes'. *Psychological Review* 107: 101–26.
- Wolf, S. 1982. 'Moral Saints'. *The Journal of Philosophy* 79: 419–39.
- Wolf, S. 2016. 'Two Concepts of Rule Utilitarianism'. In M. Timmons (ed.) *Oxford Studies in Normative Ethics, Volume 6*. Oxford: Oxford University Press, 123–44.
- Wolff, J. 2006. 'Making the World Safe for Utilitarianism'. In A. O'Hear (ed.) *Political Philosophy. Royal Institute of Philosophy Supplement*. Cambridge: Cambridge University Press, 1–22.
- Wollstonecraft, M. 1790. *A Vindication of the Rights of Men, in a Letter to the Right Honourable Edmund Burke; Occasioned by His Reflections on the Revolution in France*. London: J. Johnson.
- Woodard, C. 2008a. *Reasons, Patterns, and Cooperation*. New York: Routledge.
- Woodard, C. 2008b. 'A New Argument against Rule Consequentialism'. *Ethical Theory and Moral Practice* 11: 247–61.
- Woodard, C. 2009. 'What's Wrong with Possibilism'. *Analysis* 69: 219–26.
- Woodard, C. 2013. 'The Common Structure of Kantianism and Act Utilitarianism'. *Utilitas* 25: 246–65.
- Woodard, C. 2016. 'Hybrid Theories'. In G. Fletcher (ed.) *The Routledge Handbook of Philosophy of Well-Being*. Abingdon: Routledge, 161–74.
- Woodard, C. 2017. 'Three Conceptions of Group-Based Reasons'. *The Journal of Social Ontology* 3: 107–27.
- Wren-Lewis, S. 2014. 'How Successfully Can We Measure Well-being through Measuring Happiness?' *South African Journal of Philosophy* 33: 417–32.
- Yeo, S. L. 2017. 'Measuring the Consequences of Rules: A Reply to Smith'. *Utilitas* 29: 125–31.
- Young, I. M. 2002. 'Difference as a Resource for Democratic Communication'. In D. Estlund (ed.) *Democracy*. Malden: Blackwell, 213–33.
- Ziegler, F. V., and Tunney, R. J. 2012. 'Decisions for Others Become Less Impulsive the further away They Are on the Family Tree'. *PLoS ONE* 7: 1–5.
- Zimmerman, M. J. 1996. *The Concept of Moral Obligation*. Cambridge: Cambridge University Press.

Index

Note: Figures and tables are indicated by an italic 'f', 't' and notes are indicated by 'n.' following the page numbers.

- abhorrent actions 18, 36, 212–13
absolute prohibitions 19–20, 117–18, 161;
 see also constraints, agent-relative
act-based reasons 11, 89–93, 96–9, 101, 106,
 108–12, 122n. 24, 123n. 29, 135, 188–9,
 191–2, 204
actualism vs. possibilism 61–2, 94–5
Act Utilitarianism 18–20, 23–5, 58*f*, 59*f*,
 102n. 40, 119–20
Adams, Robert 8n. 16, 53n. 37, 56n. 40,
 201n. 27, 206n. 36, 207–8, 208n. 44
agent-neutrality and agent-relativity 5n. 10,
 20, 24, 39n. 7, 87, 95–6, 117–18, 120–1,
 123, 189, 213; *see also* constraints, agent-
 relative; *see also* options, agent-relative
aggregation of interests 17–18, 26–9, 151–2;
 see also separateness of persons
alienation test 73–81, 84
Anderson, Elizabeth 138n. 3, 148n. 26, 151n. 34
Anscombe, G. E. M. 19n. 18, 43n. 22, 46n. 27
Aristotle 43n. 22, 138n. 5, 141n. 14
Arneson, Richard 63n. 2, 69n. 19, 120n. 18,
 141n. 13, 148n. 26, 150n. 31, 153n. 37,
 183n. 48
Arpaly, Nomy 202n. 28, 30
Asimov, Isaac 169n. 19
Austin, John 34n. 61
autonomy 5, 9, 63, 66, 71, 82, 211–12

beauty 9, 16–17, 64–5, 211–12
Bentham, Jeremy 3–4, 29, 33, 113, 128n. 39,
 147, 165, 181, 206n. 37
Binmore, Ken 31n. 53, 176n. 31
Bradley, Ben 61n. 51, 63n. 1, 66n. 8, 70n. 25,
 206n. 37, 207n. 40
Brandt, Richard 23n. 33, 77n. 39, 86n. 2,
 101n. 34, 122n. 24, 123n. 30, 125n. 33,
 133n. 47, 139n. 7, 156n. 42
Brennan, Geoffrey 203n. 33
Brink, David 23n. 31, 25n. 38, 118n. 13, 119,
 120n. 18, 121n. 22–3, 122–3, 167

Broad, C. D. 209n. 45
Broome, John 15n. 6, 28n. 48, 52n. 36
Buchanan, Allen 140n. 9, 148n. 27, 175n. 29,
 178n. 34
Burch-Brown, Joanna 197n. 18
Butler, Bishop 18–19
Bykvist, Krister 52n. 34, 54n. 38, 78n. 42,
 90n. 13, 199n. 23

calculation
 pervasive calculation 33–5, 216
 role in utilitarianism 6–7, 30, 166–8, 198,
 207–8
 see also decision procedure
Carlyle, Thomas 14, 18
Chappell, Richard Yetter 22n. 28, 26,
 41n. 11–12, 42n. 20
Chesterton, G. K. 133
cluelessness 192, 196–8, 216–17; *see also*
 decision procedures
coercion 30–1, 129, 139–40, 139n. 6,
 175n. 29, 176–7, 185, 215–16
Cohen, G. A. 138n. 3, 139n. 8, 148n. 26, 150n. 31
Cohen, Joshua 180n. 38
collapse objection to Rule Consequentialism
 101–5, 108–9; *see also* incoherence
 objection to Rule Consequentialism
compensation cases 88, 90–2
consequences, visible vs. invisible 193,
 194n. 15, 197
consequentialism 4, 7–9, 15
 Act Consequentialism 4–5, 85–6, 89–90,
 96–7, 101–5, 108, 112, 189
 Collective Consequentialism 94–5
 direct vs. indirect forms 2, 4, 36n. 62,
 57–62, 102–3, 121–3, 131, 135–6, 160,
 166–7, 174–5
 Global Consequentialism 5n. 9, 59–61
 monist vs. pluralist 60–1, 89
 Motive Consequentialism 4–5
 relationship to teleology 5n. 10

- consequentialism (*cont.*)
 Rule Consequentialism 4–5, 47, 94–5, 100, 108, 111–12, 197n. 18
- consequentializing 87
- constraints, agent-relative 20–1, 24, 88, 95–6, 105–6, 109, 112, 116n. 8, 117–19, 121, 135, 141, 145, 161, 189–90, 192, 212–15
- cooperation 95n. 24, 99–100, 103–4, 106–8, 110, 111n. 55, 174–6
- Cooperative Utilitarianism 99, 104n. 43, 105
- coordination problems 94–5, 103–5, 111
- cost-benefit analysis 1
- Cottingham, John 133n. 48
- Crisp, Roger 5n.11, 16, 43n. 22, 45n. 24, 46n. 27, 50n. 31, 55n. 39, 63n. 1, 65n. 7, 141n. 13, 150n. 32, 198n. 19, 199n. 21, 209n. 46
- Cullity, Garrett 23n. 29
- Dancy, Jonathan 39nn. 4–5, 190n. 5
- decision procedures 34–6, 167, 195, 216–17
 decision rules vs. decision characteristics 195
 meta-decision rules 198–9
 validation 196–8
 vs. criterion of rightness 34–5
see also calculation
- de Lazari-Radek, Katarzyna 147n. 24
- demandingness 9–10, 21, 87, 213–14
 demands on action vs. demands on motivation 22–3
- democracy 31–2, 168, 180
 aggregative conception 168
 as constraint on sinister interests 181–2
 deliberative conception 180–1
 as educative 182–3
 and political culture 184
 as requirement of legitimacy 184
 as safeguard against incompetence 182
 utilitarian reasons for favouring 168–70, 180
- deontic status of actions 44–5, 44t; *see also* rightness
- desert 27–8, 141, 202–3
- Dickens, Charles 33
- Dietz, Alexander 99n. 29, 105, 111–12, 189–90, 213
- diminishing marginal utility of
 resources 153–4; *see also* equality,
 simple argument that utilitarianism
 favours equality
- diversity, benefits of 180–1
- Dorsey, Dale 67nn. 11–12, 16, 70n. 25, 75n. 35, 77n. 41
- Dostoevsky, Fyodor 18–20
- Driver, Julia 5n. 9, 44n. 23, 59n. 45, 200–1, 206–7
- Dworkin, Ronald 118, 118n. 12, 137n. 1, 138n. 3, 141n. 13, 148n. 26, 150n. 31
- efficiency 129n. 41, 130, 145
- eligibility of patterns 11, 92, 97, 105–8, 112, 189–90, 192, 205, 213; *see also* minimal constraint; *see also* willingness requirement
- Elster, Jon 181n. 44, 182n. 47
- embedding 61–2, 123
- epistemic manna 50n. 31
- equality 27–8, 137, 141, 146–57, 161–2
 equal importance of equal sized benefits and harms 147–8, 186
 equality of basic rights 148–50, 161
 formal justice 146–7, 151, 161
 as a goal rather than a requirement of justice 161–2
 kinds 146
 Lerner's argument 156–7
 and moral status of non-human animals 13n. 1, 148n. 26
 and non-discrimination 159–60
 simple argument that utilitarianism favours equality 154, 155f
 substantive equality 150–1
- esoteric morality 166–7; *see also* publicity
- Estlund, David 166n. 7, 175n. 29, 178n. 34
- evaluative focal points 58, 60–1
- Ewing, A. C. 8n. 17, 18nn. 14, 16, 41n. 11, 86n. 3
- expensive needs 137, 154, 156–7
 two senses distinguished 158
- Feldman, Fred 14n. 4, 27n. 42, 28n. 46, 63n. 1
- Fletcher, Guy 63n. 1, 64n. 5, 69n. 19, 95n. 20
- Foot, Philippa 1, 42n. 17, 86n. 3, 119n. 15
- Frankfurt, Harry 141n. 13, 150n. 32, 154n. 39
- Frazier, Robert 193n. 12, 196
- friendship 16, 22, 34–6, 66, 71, 188, 208–9
- fungibility of interests 26
- future people 13n. 1, 155–7, 169–70, 184, 193, 216

- Gibbard, Allan 104n. 43, 110, 120n. 18, 132n. 46
 Goodin, Robert 30, 165, 169n. 15
 goodness 4–6, 8–9, 26, 39–43
 actual vs. expected 43, 47–8, 50, 52,
 89n. 11, 156, 195, 199–200
 attributive vs. predicative 42n. 17
 overall 42–3
 simpliciter 41–3
 see also value
 Government House utilitarianism 164
 Gradgrind, Mr. 33
 Griffin, James 6n. 13, 63n. 2, 70n. 23,
 124n. 31, 148n. 27
 guise of the good 75n. 33
- Hare, R. M. 23n. 33, 120n. 18, 122n. 24,
 127, 176n. 31
 Hart, H. L. A. 115n. 4, 122nn. 25–6, 146n. 23,
 172n. 24
 Hausman, Daniel 71n. 26, 81n. 44
 Heathwood, Chris 63n. 2, 66n. 10, 69n. 21,
 71n. 27
 hedonism: *see* well-being
 Hobbes, Thomas 129n. 41, 146n. 22, 176–7
 Hohfeldian incidents 114–15, 116n. 10,
 128n. 40, 214
 Holroyd, Jules 203n. 32
 Hooker, Brad 7n. 14, 17n. 12, 21n. 25,
 22n. 27, 36n. 62, 47–53, 48f, 55–6,
 61n. 50, 63n. 1, 66n. 9, 68n. 17, 86n. 2,
 100n. 32, 101, 101n. 33, 102n. 37,
 103n. 42, 104n. 43, 132n. 45, 146n. 23,
 148n. 25, 150n. 30, 153n. 38, 191n. 7,
 197n. 18, 206n. 37
 Howard-Snyder, Frances 102n. 40, 193n. 12
 Hurka, Thomas 18n. 16, 63n. 2, 207n. 41
 Hurley, S. L. 41n.13, 69n. 19, 70n. 24, 76n. 37,
 99n. 29
- inadequate account of psychology 33, 216–17
 incoherence objection to Rule
 Consequentialism 101–3, 105; *see also*
 collapse objection to Rule
 Consequentialism
 intuitions 7–8, 69–70, 135, 152
 integrity 22–3, 34, 106, 205
- Jackson Cases 51–2, 55–6, 195, 199–200
 Jackson, Frank 43n. 21, 48n. 30, 50n. 31,
 51–2, 52n. 36, 55–6, 93n. 16, 95n. 24,
 195, 199–200
- Jim and Pedro 106–7, 110
 justice 4, 6, 11, 15n. 8, 17, 32, 133,
 137–41, 146–7, 150–1, 165n. 5,
 179–80, 208–9, 216
 distributive justice vs. other kinds of
 justice 138–9
 formal justice 146–7, 151, 161
 ‘ideally requires’ vs. ‘non-ideally
 requires’ 144–5
 procedural character 145–6, 172n. 23
 justification, doxastic vs. propositional
 51n. 33
- Kagan, Shelly 20n. 20, 22n. 28, 24nn. 36–7,
 28n. 46, 58–60, 58f, 59f, 61nn. 48–9,
 62n. 52, 64n. 3, 68n. 18, 87n. 4, 88n. 7,
 89n. 9, 95n. 25, 103n. 41
 Kagan-diagrams 58f, 59f, 60–1
 Kamm, Frances 106n. 48, 116n. 8
 Kant, Immanuel 19n. 18, 94–5
 Kantianism 21, 94–5
 Korsgaard, Christine 95n. 20
 Kraut, Richard 1, 14n. 5, 63n. 2, 67n. 14,
 69n. 20, 70n. 22
 Kymlicka, Will 148n. 26
- large numbers 28
 legitimacy 2, 12, 31–2, 139–40, 170–4, 180,
 184–5, 215–16
 as constraint on feasibility 170,
 177–8, 185
 importance in political psychology
 173–4
 as mechanism for coping with
 disagreement not resolving it 171
 normative vs. non-normative
 conceptions 175–7
 priority of established methods 171
 prized by leaders 173–4
 relationship to duty to obey 175n. 29,
 178–9
 relationship to rightness 172–3
 legitimate expectations 130–1, 143–5, 161
 Lenman, James 193–4, 197n. 18
 Lerner, Abba 154n. 39, 156–7
 levelling-down objection 152–3
 Lewis, David 67nn. 11, 13
 lexical priority 119, 161
 libertarianism 214
 liberty 10–11, 83
 Lilburne, John 182

- Locke, John 146n. 22
- Lyons, David 20n. 21, 102n. 37, 115n. 4, 122n. 24, 140n. 12
- McClenen, Edward 94–5
- McCloskey, H. J. 18, 21n. 24
- Machiavelli, Niccolò 30
- McElwee, Brian 46n. 26, 53n. 37, 55n. 39, 86n. 2, 208n. 43
- McHugh, Conor 41n. 11, 42
- Mason, Elinor 34n. 59, 52n. 34, 194n. 13
- mattering 212
- methods of normative ethics 3, 7–8, 13, 23–4, 69–70, 86, 191
- Mill, James 32n. 54, 33n. 57, 181–2, 181n. 43
- Mill, John Stuart 3n. 5, 14, 16n. 9, 32n. 54, 33, 45n. 24, 53n. 37, 57n. 41, 83n. 50, 119, 120n. 18, 121–3, 128n. 39, 138n. 3, 140n. 9, 147, 155n. 41, 166, 168n. 14, 182–3, 197–8, 211
- Miller, Dale 193
- Miller, David 171n. 21, 180n. 37
- minimal constraint 97, 111–12
- Moore, G. E. 14n. 5, 17n. 10, 50–2, 51f, 55–6, 64n. 4, 193n. 11, 195
- Motive Utilitarianism 57–8, 58f, 60–1
- Mulgan, Tim 7n. 14, 8n. 16, 13n. 1, 21n. 25, 23nn. 32, 34, 24n. 36, 46n. 26, 61nn. 48, 50, 92n. 14, 95n. 21, 100n. 32, 110n. 53, 133n. 49, 169n. 18, 184n. 55, 202–4
- Murphy, Liam 139n. 8, 146n. 22
- Muturs 206–7, 208n. 42
- Nagel, Thomas 146n. 22
- Nietzsche, Friedrich 14n. 2
- non-human animals 13n. 1, 128n. 38, 148n. 26, 169–70, 211n. 1, 216
- Nozick, Robert 19, 20n. 21, 28n. 44, 119n. 15, 137–8, 137n. 1, 140n. 10, 141, 145–6, 171n. 20
- O'Neill, Martin 150n. 33, 157n. 45
- options, agent-relative 24–5, 214
- options, range of actions facing agent 7, 35, 49, 72–3, 82–3, 90n. 13, 98–9, 129–30, 195
- 'outcome', definition of 15n. 6
- overridingness of morality 24–5, 56–7
- Owens, David 116n. 10
- Page, Scott 181n. 41
- Parfit, Derek 5n. 9, 6n. 12, 13n. 1, 17nn. 11–12, 25n. 38, 27n. 43, 28nn. 45, 47, 52n. 34, 52n. 36, 56n. 40, 59n. 45, 86, 141n. 13, 150nn. 32–3, 153n. 36, 192n. 8
- partiality 5n. 10, 22–3, 34, 188, 208–9
- pattern-based reasons 11, 90–7, 100–1, 103, 123–6, 131–2, 178, 188–92, 204, 213
- strength 106–8
- whether they conflict with act-based reasons 108–11
- patterns of action 93
- perspectival/non-perspectival pluralism 52–3, 53f, 55–6, 61n. 48, 97n. 27, 187
- Pettit, Philip 5n. 9, 59–60, 117n. 11, 118n. 14, 120nn. 18–19, 125n. 33, 128n. 40, 174n. 28, 180n. 38, 203n. 33
- pleasure 3n. 5, 9, 14–17, 33, 39, 63, 65–6, 68, 83n. 50, 120n. 19, 188–9, 206–7
- Pogge, Thomas 156n. 43
- political realism 163–4, 171n. 22, 175, 185
- politics, distinctiveness of 9–10, 12, 30–2, 163–4, 171n. 22, 185
- population ethics 6n. 12, 13n. 1, 17, 91–2, 155n. 41
- Portmore, Douglas 5n. 10, 15n. 6, 25n. 38, 39n. 7, 86n. 2, 87nn. 5–6, 90n. 13, 98n. 28
- Posner, Richard 127n. 37
- practices 109–10, 112, 122–3, 129, 203–5
- praiseworthiness and blameworthiness 10, 12, 43–4, 46–51, 48f, 51f, 53f, 53–6, 54f, 55f, 62, 186–7, 201, 207–9
- appropriateness 201, 203–5
- counterfactual 205
- rightness of praising theory 202–3
- valuable trait theory 201–2
- principle of indifference 194
- prioritarianism 17–18, 141, 147–8, 150–1, 153, 212
- promises 18–20, 87–90, 93, 97–9, 101, 104–5, 108–10, 110n. 54, 114–17, 129, 132, 140–1, 172, 192, 197, 199–200, 212
- prospection 81–2
- prudence 25, 44–5, 56–7, 69, 71–3, 75n. 34, 79, 81–3, 145, 168n. 14, 198n. 19
- publicity 32, 167n. 10; *see also* esoteric morality
- punishment 18–19, 21n. 24, 121–2, 173, 203n. 32, 212–13

- Railton, Peter 10–11, 23n. 31, 34nn. 59–60, 67–9, 73–6, 77n. 39, 80, 84, 167n. 12
- Railton's idea 10–11, 67–9, 73–6, 80, 84
- Rawls, John 1n. 4, 7n. 14, 14n. 5, 15n. 8, 20n. 21, 23n. 33, 25–9, 31n. 52, 32, 109n. 52, 119n. 17, 122n. 25, 129n. 41, 131n. 43, 133n. 49, 137–8, 139nn. 8–9, 142–7, 146n. 21, 150n. 30, 161n. 49, 163–5, 167n. 10, 171, 171n. 20, 172n. 23, 180–1, 185, 191n. 7
- Raz, Joseph 115n. 4, 116–17, 132n. 46, 140n. 11, 141n. 13, 150n. 32, 172, 175n. 29, 178n. 35
- reasonable disagreement 30–1, 164
- 'reason implies can' 90n. 13
- reasons for action 8–9, 37, 47–57
- content-dependent vs. content-independent 172
 - contributory nature 39
 - general reasons 19–20
 - interaction 7, 39–40, 190, 214
 - internalism about 41
 - normative vs. motivating 37–8
 - perspectival vs. non-perspectival 38, 43, 47–52, 89n. 11, 186–8, 192, 194, 209–11
 - relationship to rightness 39, 46–52, 56–7, 190, 214
 - teleological conception of 39–41
 - weight or strength of 39–40, 101, 103, 106–8, 111–12, 190–2, 213
 - see also* act-based reasons
 - see also* pattern-based reasons
- reflective equilibrium 7–8, 86, 87n. 6, 191
- Regan, Donald 52n. 34, 99–100, 104n. 43, 108n. 51, 110
- repugnant conclusion 27n. 43, 28n. 45
- rightness 43–52
- perspectival vs. non-perspectival 47–52, 186–8, 192, 194, 209–10
 - prudential vs. moral 44–5, 56–7
 - relationship to reasons 39, 46–52, 56–7, 190, 214
 - see also* deontic status of actions
- rights 11, 19–21, 32, 40, 110, 113, 138–46, 161, 189, 192, 196, 208–9, 211–15
- as features of best decision procedure 120–1, 208–9
 - benefits of pattern of respect for moral rights 126
 - contingency of 133–5
 - equality of basic rights 148–51
 - human 124n. 31, 126–7, 148n. 27
 - infringing vs. violating 115–16
 - legal 11, 113, 115–16, 130–1, 134–5, 143
 - nature of pattern of respect for moral rights 125–6
 - relation between moral and legal rights 113, 130–1, 134–5, 145, 161
 - relation to reasons 116–17
 - sanction theory 121–3
- Riley, Jonathan 32n. 54, 168
- Roemer, John 28n. 44, 138n. 3, 151n. 35
- Ross, Sir David 17n. 10, 18–20, 28n. 46, 141n. 14, 192
- Rule Utilitarianism 2, 11, 21, 57–8, 58f, 59–61, 133, 192, 213
- sadism 9, 16–17
- saints and sinners 27–8, 141
- Scalar Utilitarianism 45–7
- Scanlon, Thomas 15n. 8, 40n. 10, 56n. 40, 110n. 54
- Scheffler, Samuel 1n. 4, 5n. 10, 8n. 17, 20n. 20, 23n. 34, 24n. 37, 25n. 38, 86nn. 2–3, 88n. 7, 95n. 25, 188n. 1
- Schultz, Bart 33n. 58
- Schroeder, Mark 39n. 6, 41n. 15, 87n. 5
- security 10–11, 71, 82–3, 120–1, 128–32, 134, 149–50, 158, 160, 176–7
- Sen, Amartya 4n. 7, 5n. 11, 14n. 5, 15n. 7, 18n. 13, 20n. 21, 119n. 15, 126–7, 137n. 1, 138n. 3, 150n. 31, 157n. 46, 158n. 47, 183
- separateness of persons 9–10, 25, 32, 36, 214–15; *see also* aggregation of interests
- shape of a life 27, 65n. 6
- Shaw, William 86n. 2, 142n. 16, 154nn. 39–40, 206n. 37
- Sidgwick, Henry 25n. 38, 32, 34n. 61, 57n. 41, 75n. 34, 77n. 39, 81n. 44, 128n. 39, 162n. 50, 166–7, 176n. 31, 183n. 51, 202, 206n. 37, 211
- Singer, Peter 23n. 33, 147n. 24
- Smart, J. J. C. 86n. 2, 102n. 37, 188n. 2, 191n. 6, 193n. 11, 202n. 30
- Smith, Holly 52n. 34, 52n. 36, 61n. 50, 100n. 32
- Smith, Michael 5n. 9, 41n. 15, 59–60, 75n. 35
- Sobel, David 23–4, 25n. 38, 70n. 25, 76–7
- social choice theory 169
- special obligations 20–1

- Stein, Mark 157n. 46, 158, 159n. 48
- Steiner, Hillel 115n. 4, 116n. 5, 140n. 10
- suberogatory actions 44n. 23
- sufficientarianism 141, 150–1, 153
- sum-ranking 5–7, 9, 15, 17–18, 26–9, 65n. 6,
86–7, 147, 151–3, 211–12, 214–15
purely axiological objection to 151–2
see also aggregation of interests
see also separateness of persons
- supererogatory actions 22–3, 44n. 23
- Tasioulas, John 138n. 4, 140n. 10
- Team Reasoning 94–5
- technocracy 31–2, 165–8, 170, 181n. 42, 185,
215–16
- Temkin, Larry 27n. 43, 28n. 46, 29n. 49,
141n. 14, 153n. 36
- Thomson, Judith Jarvis 1n. 4, 41–3, 64n. 4,
116n. 6
- trying things 80–4
- Tyler, Tom 173–4, 176n. 32
- ‘utilitarian’, connotations 3–4
- utility monster 28n. 44
- value 5–6, 9, 14
instrumental vs. noninstrumental 65
see also goodness
- valuing 66–7, 69–70, 75–6
changing values 77, 81–2
informed vs. uninformed 76–7
possibility of incoherent values 76
see also alienation test
- Velleman, David 27n. 42, 65n. 6, 75nn. 33–4
- virtues and vices 4, 6, 12, 30, 35–6, 205,
209–10
examples of virtues 208–9
virtues and admirability 200–1, 206–9
virtue and rationality 210n. 47
virtues as excellences 207–8
- virtue ethics 21
- Way, Jonathan 41n. 11, 42
- Weber, Max 30–1, 176n. 32, 215–16
- welfarism 5–7, 9–11, 15–18, 84, 86–7, 89n. 10,
118–19, 189–90, 201, 204–5, 211–12
- well-being 1, 5–7, 9–11, 15–17, 26–8, 42, 63,
119, 186, 211–12
causes vs. constituents of 65
desire theories of 63, 66, 70–1
enumerative vs. explanatory questions
about 65–6
hedonism 14, 16n. 9, 63, 66, 68, 71, 89n. 10
hybrid theories of 68, 73
ill-being 64n. 3, 69–71
narrow vs. broad sense of 64n. 3
objective list theories of 63, 66, 71
perfectionism 63
science of 71–2
subjectivism about 10–11, 63, 66–70,
75n. 35
- Wenar, Leif 114n. 3, 115, 116n. 10
- Williams, Bernard 1n. 4, 23n. 31, 30n. 50,
32n. 56, 34n. 60, 41n. 13, 106–7,
151n. 35, 163n. 1, 166–7, 175n. 30,
176–7, 181n. 42, 188n. 1
- willingness requirement 11n. 20, 99–100,
105, 110–12, 123, 189–90
- Wolf, Susan 102nn. 38–9, 188n. 1
- Wollstonecraft, Mary 157n. 45
- wronging someone vs. acting wrongly 20–1,
113–14
- Zimmerman, Michael 62n. 53, 95n. 24