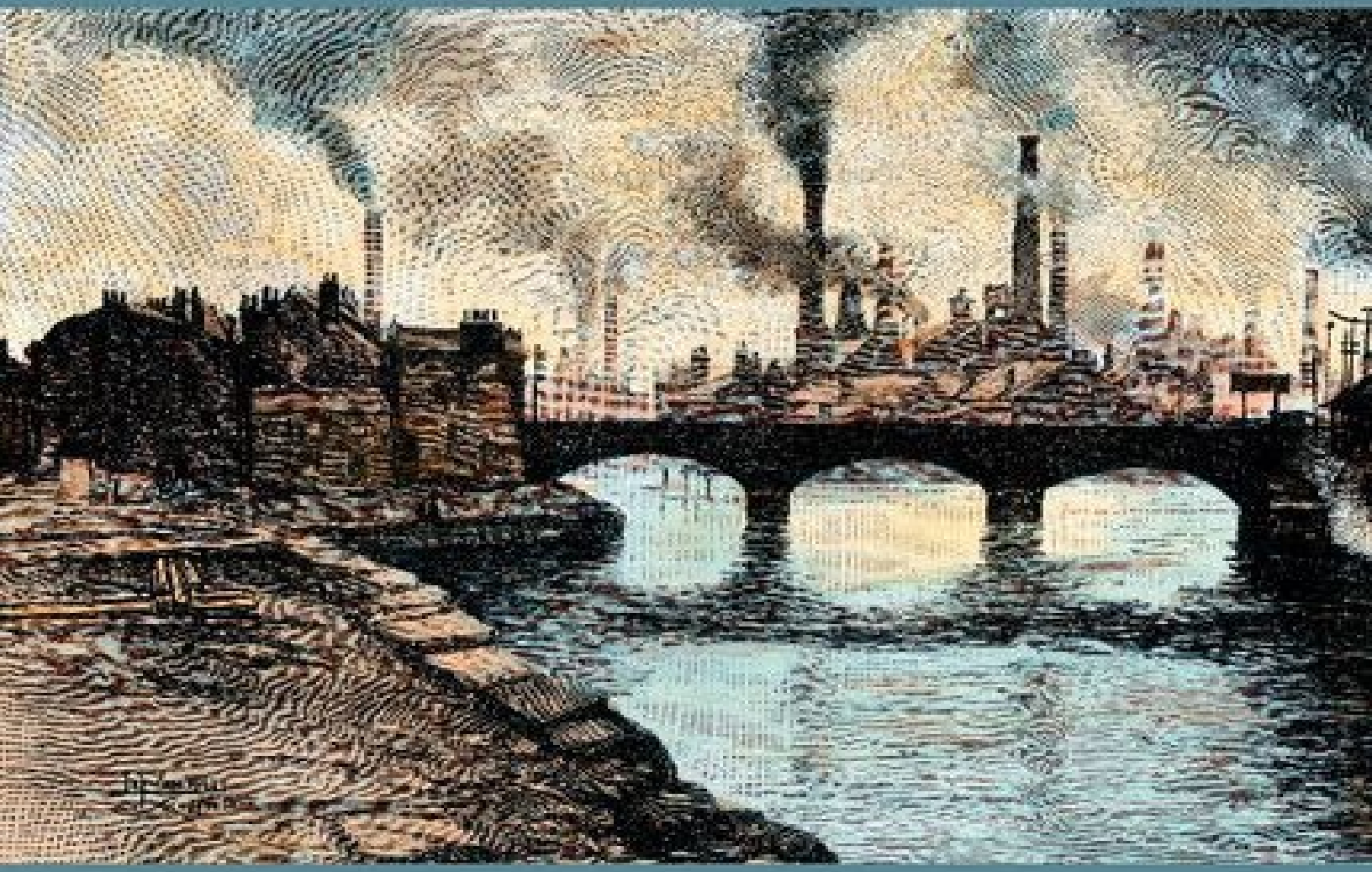
The Cambridge Companion to Utilitarianism



EDITED BY
BEN EGGLESTON
AND DALE E. MILLER

CAMBRIDGE

The Cambridge Companion to Utilitarianism

Utilitarianism, the approach to ethics based on the maximization of overall well-being, continues to have great traction in moral philosophy and political thought. This *Companion* offers a systematic exploration of its history, themes, and applications. First, it traces the origins and development of utilitarianism via the work of Jeremy Bentham, John Stuart Mill, Henry Sidgwick, and others. The volume then explores issues in the formulation of utilitarianism, including act versus rule utilitarianism, actual versus expected consequences, and objective versus subjective theories of well-being. Next, utilitarianism is positioned in relation to Kantianism and virtue ethics, and the possibility of conflict between utilitarianism and fairness is considered. Finally, the volume explores the modern relevance of utilitarianism by considering its practical implications for contemporary controversies such as military conflict and global warming. The volume will be an important resource for all those studying moral philosophy, political philosophy, political theory, and the history of ideas.

BEN EGGLESTON is Associate Professor of Philosophy at the University of Kansas. He is co-editor (with Dale E. Miller and David Weinstein) of *John Stuart Mill and the Art of Life* (2011).

DALE E. MILLER is Professor of Philosophy at Old Dominion University. He is author of *J. S. Mill: Moral, Social and Political Thought* (2010) and co-editor of *Morality, Rules, and Consequences* (with Brad Hooker and Elinor Mason, 2000) and *John Stuart Mill and the Art of Life* (with Ben Eggleston and David Weinstein, 2011).

Other volumes in the series of Cambridge Companions

For a list of titles published in the series, please see end of book.

The Cambridge Companion to Utilitarianism

Edited by

Ben Eggleston
University of Kansas

Dale E. Miller
Old Dominion University

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Published in the United States of America by Cambridge University Press, New York

Cambridge University Press is part of the University of Cambridge.
It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107656710

First published 2014

Printed in the United Kingdom by Clays, St Ives plc

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging in Publication data
The Cambridge companion to utilitarianism / edited by Ben Eggleston, University of Kansas; Dale E. Miller, Old Dominion University.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-107-02013-9 (alk. paper)
1. Utilitarianism. I. Eggleston, Ben, 1971– II. Miller, Dale E., 1966–
B843.C36 2014
171′.5–dc23
2013043752

ISBN 978-1-107-02013-9 Hardback
ISBN 978-1-107-65671-0 Paperback

Contents

Contributors

Ben Bradley is Professor and Chair of the Philosophy Department at Syracuse University. He is the author of *Well-Being and Death* (2009) and co-editor of the *Oxford Handbook of Philosophy of Death* (2012).

Krister Bykvist is Professor of Practical Philosophy at the Department of Philosophy, Stockholm University, Sweden. Before taking up this professorship he was a Fellow and Tutor in Philosophy at Jesus College, Oxford. He is the author of "No Good Fit: Why the Fitting Attitude Analysis of Value Fails" (*Mind*, 2009), "Can Unstable Preferences Provide a Stable Standard of Well-Being?" (*Economics and Philosophy*, 2010), and *Utilitarianism: A Guide for the Perplexed* (2010).

James E. Crimmins is Professor of Political Theory at Huron University College, Western University, Canada, and Fulbright Research Chair at Vanderbilt University, Tennessee, 2013–2014. His publications include *Religion, Secularization and Political Thought* (1989, reprinted 2013), *Secular Utilitarianism* (1990), *Utilitarians and Religion* (1998), *Jeremy Bentham's Auto-Icon and Related Writings* (2002), *On Bentham* (2004), *Utilitarians and Their Critics in America 1789–1914* (with Mark G. Spencer, 4 vols., 2005), *Church-of-Englandism and Its Catechism Examined* in the *Collected Works of Jeremy Bentham* (with Catherine Fuller, 2011), *Utilitarian Philosophy and Politics* (2011, paperback 2013), and *The Bloomsbury Encyclopedia of Utilitarianism* (2013).

Roger Crisp is Uehiro Fellow and Tutor in Philosophy at St. Anne's College, Oxford, and Professor of Moral Philosophy at the University of Oxford. He is the author of *Mill on Utilitarianism* (1997) and *Reasons and the Good* (2006). For several years, he edited *Utilitas*, and he is an associate editor of *Ethics*. He has edited the *Oxford Handbook of the History of Ethics* (2013).

Julia Driver is Professor of Philosophy at Washington University in St. Louis. She is the author of *Uneasy Virtue* (2001), *Ethics: The Fundamentals* (2006), and *Consequentialism* (2012), as well as articles in a variety of journals such as *Journal of Philosophy*, *Australasian Journal of Philosophy*, *Hypatia*, and *Philosophy and Phenomenological Research*. She is an associate editor of *Ethics* and a co-editor of *Journal of Ethics and Social Philosophy*.

Ben Eggleston is Associate Professor of Philosophy at the University of Kansas. He is the author of several articles on utilitarianism and related topics in journals such as *Utilitas*, *Mind*, and *Philosophical Quarterly*. He is also a co-editor, with Dale E. Miller and David Weinstein, of *John Stuart Mill and the Art of Life* (2011).

Chris Heathwood is Associate Professor of Philosophy at the University of Colorado at Boulder, where he works mainly in theoretical ethics. He has written on well-being, the nature of pleasure, and various topics in metaethics. He also has interests in metaphysics and the philosophy of religion.

Colin Heydt is Associate Professor of Philosophy at the University of South

Florida. His work focuses on the history of ethics, with special attention to seventeenth- through nineteenth-century British thought. He is the author of *Rethinking Mill's Ethics: Character and Aesthetic Education* (2006) as well as chapters in edited collections and articles published in *Journal of the History of Philosophy*, *British Journal for the History of Philosophy*, *History of Philosophy Quarterly*, and *Hume Studies*.

Brad Hooker is Professor of Philosophy at the University of Reading. He is the author of *Ideal Code, Real World: A Rule-consequentialist Theory of Morality* (2000) and a co-editor (with Dale E. Miller and Elinor Mason) of *Morality, Rules, and Consequences* (2000). His paper "Fairness" appeared in *Ethical Theory and Moral Practice* (2005) and his "Fairness, Needs, and Desert" appeared in M. H. Kramer, C. Grant, B. Colburn, and A. Hatzistavrou (eds.), *The Legacy of H. L. A. Hart* (2008).

Elinor Mason is Lecturer in Philosophy at the University of Edinburgh. She is a co-editor (with Brad Hooker and Dale E. Miller) of *Morality, Rules, and Consequences* (2000), and her work has appeared in numerous journals including *American Philosophical Quarterly*, *Ethical Theory and Moral Practice*, *Ethics*, *Philosophical Studies*, and *Utilitas*.

Dale E. Miller is Professor of Philosophy at Old Dominion University. He is the author of *J. S. Mill: Moral, Social and Political Thought* (2010). He is also a co-editor of *Morality, Rules, and Consequences* (with Brad Hooker and Elinor Mason, 2000) and *John Stuart Mill and the Art of Life* (with Ben Eggleston and David Weinstein, 2011).

Tim Mulgan is Professor of Philosophy at the University of Auckland, and Professor of Moral and Political Philosophy at the University of St. Andrews. He is the author of *The Demands of Consequentialism* (2001), *Future People* (2006), *Understanding Utilitarianism* (2007), and *Ethics for a Broken World* (2011).

Daniel C. Russell is Professor of Philosophy in the Center for the Philosophy of Freedom at the University of Arizona, and the Percy Seymour Reader in Ancient History and Philosophy at Ormond College, University of Melbourne. His research focuses on ancient and contemporary ethics. He is the author of *Plato on Pleasure and the Good Life* (2005), *Practical Intelligence and the Virtues* (2009), and *Happiness for Humans* (2012), and the editor of the *Cambridge Companion to Virtue Ethics* (2013).

William H. Shaw is Professor of Philosophy at San Jose State University. In addition to essays in various professional journals, he is the author of *Marx's Theory of History* (1980), *Moore on Right and Wrong* (1995), *Contemporary Ethics: Taking Account of Utilitarianism* (1999), *Business Ethics* (8th edn., 2013), and *Moral Issues in Business* (with Vincent Barry, 12th edn., 2012). He has edited or co-edited six books, including *G. E. Moore's Ethics* (2005), *Philosophy of Law* (5th edn., 2009), and *Social and Personal Ethics* (7th edn., 2012).

Jens Timmermann is Reader in Moral Philosophy at the University of St.

Andrews. He is the author of *Sittengesetz und Freiheit* (2003) and of *Kant's "Groundwork of the Metaphysics of Morals": A Commentary* (2007). He is the editor of, inter alia, *Kant's "Groundwork of the Metaphysics of Morals": A Critical Guide* (2009), *Kant's "Critique of Practical Reason": A Critical Guide* (jointly with Andrews Reath, 2010), and the first German–English edition of *Kant's "Groundwork"* (2011).

Henry R. West is Emeritus Professor of Philosophy at Macalester College, USA. His recent works on John Stuart Mill and utilitarianism include *An Introduction to Mill's Utilitarian Ethics* (2004); editor, *Blackwell Guide to Mill's Utilitarianism* (2006); *Mill's Utilitarianism: A Reader's Guide* (2007); "John Stuart Mill [Addendum]," in *Encyclopedia of Philosophy*, (2nd edn., 2006); "Mill's Case for Liberty," in *Mill's On Liberty: A Critical Guide* (2008); "John Stuart Mill," in the *Routledge Companion to Ethics* (2010); "Mill and Rawls," in *Mill on Justice* (2012); "J. S. Mill," in the *Oxford Handbook of the History of Ethics* (2013); and "Utilitarianism," in the *International Encyclopedia of Ethics* (2013).

Acknowledgments

We would like to thank all of the authors who have written chapters for this book. We appreciate not only their initial contributions, which were very skillfully done, but also their patience with an intensive editing process through which we tried to make the final manuscript as clear and accessible as possible. We would also like to thank Hilary Gaskin and Anna Lowe, of Cambridge University Press, for their support of this book and their guidance through the production of it. Finally, we would like to thank Monica Shafii, a student at the University of Kansas, for her assistance with the checking and compilation of the bibliography entries; Robert Vinten, for his assistance with the compilation of the index; the College of Arts and Letters at Old Dominion University, for subsidizing Vinten's work; and Martin Barr, for his conscientious copy-editing.

Introduction

Ben Eggleston and Dale E. Miller

Utilitarianism's place in moral philosophy

It is well known that utilitarianism – the moral theory based on the maximization of overall well-being – is one of the leading theories in recent and contemporary moral philosophy. The same can be said, of course, about several moral theories. Utilitarianism, however, arguably has the distinction of being the moral theory that, more than any other, shapes the discipline of moral philosophy and forms the background against which rival theories are imagined, refined, and articulated.

At times, utilitarianism's preeminence has been evinced in the remarks of those who would most fervently wish it gone. In the middle part of the twentieth century, for example, John Plamenatz wrote that "Utilitarianism is destroyed," with "no part of it left standing."[1] In 1973, Bernard Williams concluded his "A Critique of Utilitarianism" with the following assurance to his utilitarianism-weary readers: "The important issues that utilitarianism raises should be discussed in contexts more rewarding than that of utilitarianism itself. The day cannot be too far off in which we hear no more of it."[2] Finally, in 2011, Ronald Dworkin claimed that although the rise of utilitarianism in the nineteenth century had given it ascendancy over the rights-based doctrines that defined the morality of the Enlightenment, "Now the wheel is turning again: utilitarianism is giving way once again to a recognition of individual rights."[3]

It will be interesting to see whether the passage of time is kinder to Dworkin's assessment than it has been to those of Plamenatz and Williams. Meanwhile, the remarks of other critics of utilitarianism attest to the supremacy it has enjoyed in the discipline of moral philosophy. John Rawls, for example, wrote in the preface to *A Theory of Justice* (1971) that "During much of modern moral philosophy the predominant systematic theory has been some form of utilitarianism."[4] He added that moral philosophers who did not subscribe to utilitarianism tended not to construct opposing theoretical frameworks, but to start with utilitarianism and then propose modifications of it to allay their particular concerns. What follows, Rawls concludes, is that "Most likely we finally settle upon a variant of the utility principle circumscribed and restricted in certain ad hoc ways by intuitionistic constraints."[5] In effect, the predominance of utilitarianism was so thorough as to result in a paucity of viable alternatives.

Of course, any mention of Rawls in this context must acknowledge that his treatise itself immediately reinvigorated and profoundly reshaped the discipline of moral philosophy, giving new energy, sophistication, and contemporary relevance to social-

contract and Kantian ways of thinking. But utilitarianism was not dislodged from its place at the core of moral philosophy. In 1982, T. M. Scanlon, a fellow contractualist whose views were thus much closer to Rawls's than to any form of utilitarianism, wrote the following:

> Utilitarianism occupies a central place in the moral philosophy of our time. It is not the view which most people hold; certainly there are very few who would claim to be act utilitarians. But for a much wider range of people it is the view towards which they find themselves pressed when they try to give a theoretical account of their moral beliefs. Within moral philosophy it represents a position one must struggle against if one wishes to avoid it.[6]

Subsequently, in his *Contemporary Political Philosophy*, Will Kymlicka wrote that "Rawls believes, rightly I think, that in our society utilitarianism operates as a kind of tacit background assumption against which other theories have to assert and defend themselves."[7] Like Rawls and Scanlon, Kymlicka was writing as a critic of utilitarianism rather than as its champion. Thus, even authors who doubted the adequacy of utilitarianism nonetheless affirmed its centrality. "[U]tilitarianism tends to haunt even those of us who will not believe in it," Philippa Foot pithily wrote.[8]

What accounts for utilitarianism's persistent influence? Again, perhaps the most credible evidence comes from utilitarianism's most prominent critics. Although Rawls denies that moral rightness is based on the promotion of good consequences, he adds that he does not mean to suggest that a theory of moral rightness can ignore consequences: "All ethical doctrines worth our attention take consequences into account when judging rightness. One which did not would be simply irrational, crazy."[9] Utilitarianism's focus on consequences is also cited by Samuel Scheffler as a reason for the persistence of its influence:

> I believe that utilitarianism refuses to fade from the scene in large part because, as the most familiar consequentialist theory, it is the major recognized normative theory incorporating the deeply plausible-sounding feature that one may always do what would lead to the best available outcome overall. Despite all of utilitarianism's faults (including, no doubt, its misidentification of the best outcomes), its incorporation of this one plausible feature is in my opinion responsible for its persistence.[10]

Finally, Scanlon credits utilitarianism with having a particularly simple and compelling account of why a person might feel motivated to attend to the ideals and requirements of morality:

> In our own time, the leading substantive account of moral motivation has been that

> offered by utilitarianism. In fact it seems to me that a large part of the appeal of utilitarianism lies in the fact that it identifies, in the idea of "the greatest happiness," a substantive value which seems at the same time to be clearly connected to the content of morality and, when looked at from outside morality, to be something which is of obvious importance and value, capable of explaining the great importance that morality claims for itself.[11]

The aspects of utilitarianism highlighted by these remarks are vital contributors to utilitarianism's persistent influence and its place in contemporary moral philosophy. A broader perspective is provided by a brief historical overview of the development and reception of the view.

Historical overview

The history of utilitarianism is surveyed by the first five chapters of this volume, so here a cursory summary will suffice. Fundamental elements of utilitarianism have been focal points of philosophical discourse since ancient times; the fourth- and third-century BCE philosopher Epicurus, for example, is best known for claiming that one's primary concerns should be the attainment of pleasure and, especially, the avoidance of pain. For nearly two millennia these and other foundational notions simmered as topics of philosophical discussion, but starting in the seventeenth century these ideas and related ones were embraced by a series of writers, mostly British, who assembled them with increasing sophistication into formidable philosophical systems. The most notable such writers, and their most notable works, are Jeremy Bentham (*An Introduction to the Principles of Morals and Legislation*, 1789), John Stuart Mill (*Utilitarianism*, 1861), and Henry Sidgwick (*The Methods of Ethics*, seven editions from 1874 to 1907).

As mentioned above, and as suggested by the dates just mentioned, utilitarianism surged to unprecedented prominence in the nineteenth century: sufficient time had passed for the teachings of Bentham to spread widely, and Mill was a public intellectual who enjoyed a wide readership for his prolific writings. This new prominence for the theory, however, was accompanied by a corresponding degree of opposition and criticism. In fact, several of the leading lights of the nineteenth century condemned utilitarianism publicly and vigorously.

The conservative historian and social critic Thomas Carlyle, despite his friendship with Mill, exemplified this trend. In one of a series of six lectures he gave in May 1840, Carlyle praised the prophet Muhammad's conscientious and absolutist sense of duty and said it "might put some of *us* to shame":[12]

> Benthamee Utility, virtue by Profit and Loss; reducing this God's-world to a dead brute Steam-engine, the infinite celestial Soul of Man to a kind of Hay-balance for

> weighing hay and thistles on, pleasures and pains on: – If you ask me which gives, Mahomet or they, the beggarlier and falser view of Man and his Destinies in this Universe, I will answer, it is not Mahomet![13]

Mill was outraged by this comparison. In a rare public display of anger, he rose from his seat and shouted "No!"[14] Three lectures later, Carlyle mentioned his previous disparagement of Bentham's views and affirmed it as his "deliberate opinion."[15]

Also in the 1840s, utilitarianism was known and rejected in literary circles. The historian William O. Aydelotte writes that the authors he regards as "the four most important social novelists of the decade" – Charles Dickens, Charles Kingsley, Benjamin Disraeli, and Elizabeth Gaskell – each "repudiated rationalistic utilitarianism, often through the mouth of a principal character speaking obviously for the author."[16] The case of Dickens is especially notable because of that author's lasting popularity and the intensity of his reaction to utilitarianism; one Dickens scholar reports that "he was a life-long opponent of utilitarian ideas as he understood them" and that his "fury never abated."[17] Dickens's 1854 novel *Hard Times*, with its cold-hearted, fact-obsessed Mr. Gradgrind, is often read as a denunciation of utilitarianism.[18]

In the latter half of the nineteenth century, utilitarianism was the object of severe and specific criticisms from two of the most celebrated philosophers of that era, Karl Marx and Friedrich Nietzsche. In volume I of *Capital* (1867), Marx claimed that Bentham, in particular, was in the grip of a conception of human well-being that was tied to a specific cultural context and was therefore unsuitable for a moral theory that aspired to universal applicability:

> he that would criticise all human acts, movements, relations, etc., by the principle of utility, must first deal with human nature in general, and then with human nature as modified in each historical epoch. Bentham makes short work of it. With the driest naïveté he takes the modern shopkeeper, especially the English shopkeeper, as the normal man . . . This yard-measure, then, he applies to past, present, and future.[19]

Nor did Marx credit later utilitarians with salvaging their theory to any appreciable extent: according to Peter Singer, "Marx was as scornful of utilitarianism as of any other ethical theory."[20]

Nietzsche's criticisms, though equally pointed, were "complex and varied,"[21] ranging from personal digs at Bentham[22] and Mill[23] to analytical insights about the intellectual milieu in which utilitarianism thrived. He observed, for example, that the universal benevolence elevated by utilitarianism into an ethical first principle did not seem to be manifest in the personal motives of many of the advocates of utilitarianism.[24] Perhaps

Nietzsche's most focused criticism of utilitarianism was his claim that well-being is not remotely well-suited to serve as a fundamental value.[25] In *Beyond Good and Evil* (1886), he declared that "Well-being as you understand it – that is no goal; it looks to us like an *end*! – a condition that immediately renders people ridiculous and despicable – that makes their decline into something *desirable*!"[26] Nietzsche's objection was that the state of well-being, as utilitarians understood it, was indifferent or inimical to the personal flourishing that he saw as the apotheosis of human development, since such personal flourishing often requires confronting and overcoming pain and other difficulties – or even essentially involves such states – rather than being best served by steering clear of them: "The discipline of suffering, of *great* suffering – don't you know that *this* discipline has been the sole cause of every enhancement in humanity so far?"[27] Nietzsche also argued that the development of flourishing individuals was likely to be further thwarted by another aspect of utilitarianism, specifically, the egalitarianism implicit in its concern with the *general* well-being:

> "general welfare" is no ideal, no goal, not a concept that can somehow be grasped, but only an emetic . . . the requirement that there be a single morality for everyone is harmful precisely to the higher men; in short . . . there is an *order of rank* between people, and between moralities as well. They are a modest and thoroughly mediocre type of person, these utilitarian Englishmen.[28]

Finally, Nietzsche's best-known rejoinder to utilitarianism is his remark, contained in one of the "Arrows and Epigrams" at the beginning of his *Twilight of the Idols* (1889), that "People *don't* strive for happiness; only the English do."[29] In a study of Nietzsche's moral views, Frank Cameron writes that "This emphasis on the 'perfect man' or higher type is, I believe, the central motivation underlying his critique of utilitarianism . . . Nietzsche often contrasts the 'man of utility' with the 'exemplary individual'."[30]

The twentieth century began with the publication of another book that quickly earned a place on the top shelf of utilitarian studies – G. E. Moore's *Principia Ethica* (1903). But since then, no single work has been published which appears, now, to merit a place alongside the landmark texts of Bentham, Mill, Sidgwick, and Moore. This is not to suggest that progress in the development of utilitarianism had ground to a halt, however. If the contributions of subsequent writers have been modest and incremental compared to those of earlier theorists, their accomplishments are still impressive in virtue of the many more hands that have taken up the work and their collective impact. Moral philosophers with an interest in the theory have explored the characteristics and merits of a variety of its forms, by pursuing conceptual possibilities lying along several distinct dimensions. Three of these lines of development warrant particular mention here.

First, there has been considerable evolution in utilitarianism's conception of the good to be promoted, i.e., well-being. One manifestation of this evolution has been a shift in

the meaning of 'utility', which utilitarians today frequently use as a synonym for 'well-being'. This shift arguably began in the nineteenth century, but was certainly commonly seen by the middle of the twentieth.[31] To some extent the roots of these efforts toward reconceiving well-being can be traced as far back as Mill; indeed the aspects of utilitarianism to which Dickens objected were, according to Richard Arneson, also aspects of Benthamite utilitarianism to which Mill objected.[32] And perhaps Mill's most distinctive innovation, in his thinking about utilitarianism, was his more sophisticated account of kinds of pleasure, based on a more sophisticated conception of human nature. But even formulated in that way, utilitarianism was still concerned mainly with the promotion of pleasure, prompting objections of the kind lodged by Marx and Nietzsche. As late as 1936, R. F. Harrod wrote that "The Utilitarians attempted a great generalisation and affirmed that the sole ultimate end is pleasure. It is not clear that they were successful."[33] But, in 1959, Richard Brandt wrote that "Many 'refutations' of utilitarianisms are aimed at the hedonistic features, and do not touch the utilitarianism at all,"[34] reflecting the wider range of possible conceptions of well-being that utilitarian theorists were exploring. Some of these theorists, adhering to the traditional utilitarian idea that a person's well-being depends on what that person finds appealing (in some sense), suggest a relatively slight shift in focus, from pleasure to the subtly different good of having one's desires satisfied. But others, going farther afield, claim that some things are good for people regardless of whether they find them appealing or not – typical examples include achievement, knowledge, friendship, and freedom. This profusion of theories of well-being has made utilitarianism less vulnerable to objections of the kind lodged by Marx and Nietzsche, since these objections rely largely on certain assumptions about what specific sort of human existence is recommended by utilitarianism. To be sure, some contemporary critics of utilitarianism still censure it in such terms; the influential legal theorist Richard Posner, for example, writes that "utilitarianism is a hedonistic, unsocial ethic."[35] Nevertheless, well-being is a topic on which decades of gradual progress have resulted in a markedly greater general understanding of the possible forms of utilitarianism and their relative merits.

A second important area of progress in utilitarian thought is the emergence of consequentialism as a distinct focal point of ethical theorizing. Consequentialism is the idea that morality should be based on the maximization of the good (possibly well-being, possibly something else). Thus, consequentialism is more general than utilitarianism, so utilitarianism is, in effect, a family of views within the larger consequentialist family of views. Since utilitarianism is distinguished from other forms of consequentialism by the claim that the good to be promoted is well-being, consequentialism can be understood as utilitarianism minus that claim. With consequentialism understood in this way, the emergence of it as a topic of inquiry distinct from utilitarianism can be seen as a logical extension of the thorough exploration of possible conceptions of well-being discussed in the previous paragraph: just as greater openness to different ideas about what constitutes well-being makes utilitarianism less vulnerable to objections concerned with the sort of

human existence it recommends, the emergence of consequentialism as a distinct topic of discussion brackets such objections by removing the topic of well-being from the discussion altogether. In effect, it invites those who hold such objections to assert whatever conception of the good they find appealing (whether it involves well-being or not), and just plug it in to the consequentialist framework. Of course, this strategy comes at the potential cost, for utilitarianism, of opening the door to greater consideration of non-utilitarian forms of consequentialism. But this strategy has also arguably benefitted utilitarian thought because any element of a consequentialist theory that is independent of its conception of the good can, in principle, also be an element of a corresponding form of utilitarianism, and this means that proposals, analyses, and evaluations of various forms of consequentialism can, correspondingly, lead to progress in the development and assessment of various forms of utilitarianism.

This trend is reflected in the fact that for several of the chapters in this volume, the focus of discussion is some aspect of consequentialism rather than some aspect of utilitarianism. Conversely, several other chapters make claims about utilitarianism that could easily be applied to consequentialism. In fact, when this volume was being planned, serious thought was given to its ultimately being *The Cambridge Companion to Consequentialism*. It was decided, though, to keep the emphasis of the volume on the historically most significant strand of consequentialism insofar as was practicable, while acknowledging that consequentialism, rather than utilitarianism, is the primary context in which certain contemporary topics and issues are discussed.

A third area of progress in the development of utilitarian thought is the fruitful exploration of different answers to the question of exactly how, in principle, the rightness or wrongness of an act is related to the promotion of well-being. One obvious possibility is the view, associated with act utilitarianism, that the rightness (or wrongness) of an act depends simply on its effects on well-being. But alternative possibilities have been developed as well. For example, proponents of rule utilitarianism hold that the rightness of an act depends not on its particular consequences, but on its conformity to certain rules whose moral significance, in turn, depends on their promotion of well-being. The debate between act utilitarians and rule utilitarians has been a major topic in utilitarian thought for more than half a century, and since this debate is conceptually independent of utilitarianism's claim that the good to be promoted is well-being, the same issues are explored in the equally lively debate between act consequentialists and rule consequentialists. Like the topics of inquiry discussed in the previous two paragraphs, this topic has been an area of incremental progress to which many thinkers have made valuable contributions. But contemporary thinking about these issues is most indebted to the late-twentieth-century rule-utilitarian work of Richard Brandt and the subsequent rule-consequentialist work of Brad Hooker.

These reflections on the continued and multifaceted development of utilitarianism help to explain its contemporary standing as the moral theory against which other moral theories must, of necessity, be defined and contrasted. When utilitarianism is described in

this way, it is easy to think of an unchanging monolith that has loomed so persistently mainly because of inertia. But inertia is not the whole story. For utilitarianism not only elaborates a basic insight about the moral importance of the consequences of the acts that people perform; it continues to evolve in response to new thinking about human nature and the nature of well-being, the role of rules in morality, and other ethical concerns. Clearly, it has not evolved to the point where moral philosophers are universally content to endorse it rather than pursue other possibilities. But it remains not only a venerable and preeminent, but also a vibrant and flexible, theoretical framework within moral philosophy.

Overview of the volume

While the chapters of this collection are not divided into sections, they are arranged in a logical order and by and large they fall naturally into several groups.

The first and most extensive of these groups comprises the first five chapters, which outline the history of utilitarianism from its pre-Benthamite roots into the twentieth century. Colin Heydt shows that although today we think of utilitarianism as a secular moral and political theory, much of its "pre-history" is to be found in a strand of Anglican natural law theory according to which our fundamental moral obligation is to obey God's commands. We are obligated to promote happiness, according to the Anglican utilitarians, only because God wills it, which we know because we know that he loves us. Heydt also discusses the influence on utilitarianism of some thinkers who are not utilitarians themselves, such as John Locke, Francis Hutcheson, and David Hume. Utilitarianism came into its own in the nineteenth century, and it is convenient for our purposes to see this century as being divided into three distinct periods, with one particular figure occupying a position of preeminence among utilitarian thinkers during each. James E. Crimmins picks up the story at the beginning of the century, where Jeremy Bentham establishes himself as the father of modern utilitarianism. Crimmins traces the globe-spanning impact of Bentham's work, which might almost be said to have been felt later in Britain than anywhere else. Henry R. West carries the discussion into the mid-nineteenth century, where John Stuart Mill comes to the fore. His father James – a contemporary, and for many years a close associate, of Bentham's – intended Mill from birth to become a public champion of utilitarianism. And so he did, but West shows that Mill fashions a utilitarian account of morality and justice that is distinctively his own – one that says, for instance, that lesser quantities of certain superior pleasures can be more valuable and contribute more to well-being than greater quantities of inferior pleasures. Roger Crisp shifts the focus to the late nineteenth century, when Henry Sidgwick was the leading utilitarian theorist. Crisp demonstrates that with Sidgwick utilitarianism makes major advances in rigor and sophistication. Sidgwick confronts some of the main competitors to a utilitarian theory of morality, egoism and intuitionism in particular, more systematically and explicitly than earlier thinkers had done. This

culminates with Sidgwick's claim that utilitarianism can account for whatever is attractive in "dogmatic intuitionism" and his confession that he cannot show that it is any less reasonable for us to promote our own happiness than to promote the greatest overall happiness. In the twentieth century, far too much work was done on utilitarianism for one chapter to encapsulate it all, and indeed most chapters in this companion address twentieth-century developments in specific areas. Krister Bykvist's chapter, though, considers some of the most important advances in utilitarian thought in the twentieth century that are not discussed elsewhere, in particular the arguments for the view associated with John Harsanyi and R. M. Hare. Bykvist shows that these arguments reflect a more general interest among many twentieth-century utilitarians in developing the theory with greater precision and greater use of formal or at least technical methods. In Harsanyi's case, this means the methods of welfare economics; in Hare's, it means methods drawn from the philosophy of language.

The next four chapters examine different ways of formulating the "moral standard" found within utilitarian moral theories, that is, the criterion by which morally right and morally wrong actions are distinguished. The first two of these chapters concentrate on two accounts of how utilitarian considerations are to be brought to bear on the evaluation of actions. The act-utilitarian approach, the subject of Ben Eggleston's chapter, says that an action would be right if there were nothing else that the agent could do instead that would yield more overall well-being, and otherwise would be wrong. Eggleston discusses various specific versions of this basic approach and some of the leading arguments for it. He also shows how a sophisticated "indirect" act utilitarianism, one that tells agents to follow a "decision procedure" that closely resembles our ordinary morality in many respects, might be able to meet many of the objections that have been raised against the approach. Dale E. Miller discusses an analogous cluster of issues in his chapter on rule utilitarianism, the traditional rival to act utilitarianism within the utilitarian tradition. Rule utilitarians say that whether actions are right or wrong depends on whether they are permitted or forbidden by an authoritative "moral code" or set of moral rules. What is distinctive about rule utilitarianism, vis-à-vis other moral theories with rule-based moral standards, is the role that utilitarian considerations play in determining the contents of the authoritative code. A rule utilitarian might say, for example, that a society's authoritative code is the one whose acceptance by most or all of its members would result in a higher level of overall well-being than their acceptance of any other code. Julia Driver's chapter discusses a further form of utilitarianism that has recently assumed an important place in the literature, namely, "global utilitarianism." Global utilitarians apply utilitarian considerations directly to every class of evaluands (that is, the subjects of evaluation). So they judge acts like act utilitarians, social rules like rule utilitarians, character traits like "virtue utilitarians" (who would say that a character trait is a virtue if a person's possession of it would yield more overall well-being than her possession of any alternative trait), and so on. As Driver notes, global utilitarianism allows for a nuanced form of ethical evaluation – one that lets us say, for instance, that a wrong action was produced by a virtuous disposition – while remaining thoroughly utilitarian. Finally, Elinor

Mason's chapter takes up a rather different question of formulation, one that arises due to the fact that it is usually impossible to foresee all of the consequences of any evaluand perfectly. The question takes this form for an act utilitarian: Is the right action the one that will *actually* yield the most overall well-being, the one that *the agent believes* will yield the most well-being, or the one that, roughly put, strikes the *most reasonable balance* between risk and good consequences? Objective act utilitarians will opt for the first view, subjective act utilitarians the second, and prospective act utilitarians the third. Utilitarians of every stripe – rule, global, etc. – face some version of this question.

Chapters 10 and 11, by Chris Heathwood and Ben Bradley respectively, critically examine different conceptions of well-being. Derek Parfit asserts that there are three basic categories of these:

> On *Hedonistic Theories*, what would be best for someone is what would make his life happiest. On *Desire-Fulfilment Theories*, what would be best for someone is what, throughout his life, would best fulfil his desires. On *Objective List Theories*, certain things are good or bad for us, whether or not we want to have the good things, or to avoid the bad things.[36]

Heathwood's chapter discusses desire-fulfillment conceptions of well-being and subjective conceptions of well-being more generally. (A subjective conception of well-being, roughly put, says that whether something makes a person's life go better for her depends on her attitude toward it.) Bradley's chapter looks at objective conceptions. Far from its being the case that hedonism is ignored, however, it is discussed in both chapters. As Heathwood and Bradley point out, there seem to be two explanations for why pleasure or happiness might be the sole constituent of well-being. One is that it has some unique position relative to our desires or pro-attitudes, e.g., it is the only thing that we desire for its own sake. The other is that it is good for us irrespective of our desires or attitudes. In the first case, hedonism is simply a particular subjective conception of well-being, and in the second, it is simply a particular objective conception. Elevating it into a third distinct category of conceptions, as does Parfit (in company with many other writers), therefore seems to gloss over important differences between the very different paths by which one might arrive at a hedonistic view.

The next pair of chapters are both concerned with utilitarianism's relation to other traditions in moral philosophy. In Chapter 12, Jens Timmermann explores points of similarity and difference between utilitarian and Kantian approaches. There are more of the former than might be expected, Timmermann shows: not only do both approaches seek to ground morality on a single ultimate principle and to show that their respective candidates for this principle have exerted an unmarked influence on ordinary moral thinking, but Kant regards the promotion of happiness as a duty. Yet Timmermann also establishes that the Kantian approach differs from the utilitarian one in fundamental ways and that attempts to assimilate the former into the latter – like those of Mill and R. M.

Hare – are misguided. Daniel C. Russell, in turn, shows that there are also fundamental differences between utilitarianism and the approach to moral philosophy commonly known as "virtue ethics." Russell also shows, though, that the philosophers in the virtue ethics tradition frequently overlook the importance of the consequences of actions, policies, etc. A virtuous individual would often need to incorporate cost–benefit reasoning into her deliberations about what to do, Russell argues, although she would also need to recognize situations in which this sort of reasoning is beside the point.

Brad Hooker's chapter is in a class by itself, inasmuch as it is the only chapter devoted to an extended discussion of a particular objection to utilitarianism, namely, that it is insensitive to considerations of fairness. How far this is true, Hooker shows, depends on precisely how fairness is conceived, and he surveys a variety of competing conceptions. So too does it depend on what version of utilitarianism is being considered. Hooker concludes that there is inevitably significant tension between common conceptions of fairness and act utilitarianism, but much less tension (albeit still some) between fairness and rule utilitarianism.

The final two chapters involve the application of utilitarian reasoning to questions of practical ethics that are among the most urgent such questions that confront us in the twenty-first century. William H. Shaw asks what utilitarianism has to teach us about the morality of war, including both *jus ad bellum* (the morality of going to war) and *jus in bellum* (the morality of the conduct of war). He argues that even though utilitarianism's implications about the moral permissibility of the use of force might differ somewhat from those of traditional just war theory, utilitarianism would still endorse the use of just war theory – including its absolute prohibition on the intentional targeting of civilians – as a decision procedure by political and military leaders. Tim Mulgan concludes the volume by exploring our obligations to future people from a utilitarian perspective. As he shows, this is an area in which utilitarianism may enjoy a considerable advantage over rival moral theories, in virtue of the fact that it is able to guide our behavior when we face choices that will affect which people will exist in the future just as well as when we face any other sort of choice. Still, different versions of utilitarianism will have different implications for our obligations to future people, and Mulgan considers the relative merits of, for example, versions that say that it is the total amount of well-being that is to be maximized and those that say instead that it is the average amount. He also discusses how phenomena such as climate change might force us to rethink many of the long-held assumptions on which traditional accounts of intergenerational justice have been based.

Notes

1. Cited in Scarre, *Utilitarianism*, p. 2.

2. B. Williams, "A Critique of Utilitarianism," p. 150.

3. Dworkin, *Justice for Hedgehogs*, p. 414.

4. Rawls, *A Theory of Justice*, p. vii.

5. Rawls, *A Theory of Justice*, p. viii.

6. Scanlon, "Contractualism and Utilitarianism," p. 103.

7. Kymlicka, *Contemporary Political Philosophy*, p. 10.

8. Foot, "Utilitarianism and the Virtues," p. 196.

9. Rawls, *A Theory of Justice*, p. 30.

10. Scheffler, *The Rejection of Consequentialism*, p. 4.

11. Scanlon, *What We Owe to Each Other*, p. 151.

12. Carlyle, *On Heroes, Hero-Worship, and the Heroic in History*, p. 65.

13. Carlyle, *On Heroes, Hero-Worship, and the Heroic in History*, p. 65.

14. Packe, *The Life of John Stuart Mill*, pp. 264–265; and R. Cohen, "Can You Forgive Him?" p. 60.

15. Carlyle, *On Heroes, Hero-Worship, and the Heroic in History*, p. 148.

16. Aydelotte, "The England of Marx and Mill as Reflected in Fiction," p. 43 and p. 45.

17. G. Smith, "Utilitarianism," p. 582.

18. For an overview of the literature on the attitudes toward Benthamite utilitarianism reflected in *Hard Times*, see Stone, "Dickens, Bentham, and the Fictions of the Law," p. 126, n. 2.

19. Marx, *Capital*, vol. I, p. 605, n. 2.

20. Singer, *Marx*, p. 63.

21. Strong, *Friedrich Nietzsche and the Politics of Transformation*, p. 93.

22. Nietzsche, *Beyond Good and Evil*, section 228 (p. 119).

23. Nietzsche, *Beyond Good and Evil*, section 253 (p. 144).

24. Anomaly, "Nietzsche's Critique of Utilitarianism," section 3 (pp. 5–7).

25. Anomaly disputes the centrality of these concerns in Nietzsche's critique of utilitarianism; see his "Nietzsche's Critique of Utilitarianism," section 4 (pp. 8–10).

26. Nietzsche, *Beyond Good and Evil*, section 225 (p. 116).

27. Nietzsche, *Beyond Good and Evil*, section 225 (pp. 116–117).

28. Nietzsche, *Beyond Good and Evil*, section 228 (p. 119).

29. Nietzsche, *Twilight of the Idols*, "Arrows and Maxims," section 12 (p. 157).

30. Cameron, *Nietzsche and the 'Problem' of Morality*, pp. 133–134.

31. On this topic, see the exchange between John Broome and Amartya Sen in Broome, "'Utility'"; Sen, "Utility: Ideas and Terminology"; and Broome, "A Reply to Sen."

32. Arneson, "Benthamite Utilitarianism and *Hard Times*," pp. 62–67.

33. Harrod, "Utilitarianism Revised," p. 146.

34. Brandt, *Ethical Theory*, p. 381, n. 1.

35. Posner, *The Problems of Jurisprudence*, p. 391.

36. Parfit, *Reasons and Persons*, p. 493.

1 Utilitarianism before Bentham

Colin Heydt

Introduction

This chapter examines the history of utilitarianism in early modern Britain and, more briefly, France. Utilitarianism offers one problem and two advantages as a subject of historical inquiry. The problem is terminological. "Utilitarianism" is a mid-nineteenth-century term and one that referred to a reforming legal, social, and political movement of the late-eighteenth and nineteenth centuries.[1] I will anachronistically, though in accord with present-day usage, employ the term throughout this chapter to refer to utilitarian moral thought in general. Now the advantages: First, utilitarian theory began. More precisely, we can locate its origins in England during the period 1660s–1730s. For a major theory in philosophy, that is pretty specific. Second, the beginning is recent enough to provide ample documentation of the thoughts and circumstances of early proponents of the theory. These advantages enable us to think more cogently about how to respond to important questions: What motivated people to develop and defend utilitarian ideas? In choosing to defend utilitarianism, what alternatives were these thinkers rejecting? This chapter addresses these questions in the hope of making utilitarianism more intelligible – not simply as embodying philosophical theses and arguments, but as expressing and shaping modes of moral, political, and religious life.

Utilitarianism in Britain

The history of utilitarianism in Britain prior to the late-eighteenth-century work of Jeremy Bentham is dominated by what I call "Anglican utilitarianism."[2] Its notable proponents included the eighteenth-century thinkers John Gay, John Brown, Soame Jenyns, Edmund Law, Abraham Tucker, Thomas Rutherforth, and William Paley.[3] These thinkers argue for two key theses of standard utilitarianism: all things – knowledge, virtue, health – are valuable only insofar as they promote pleasure or decrease pain, and actions are right or wrong depending on their consequences for the public good, i.e., the greatest happiness. While the first thesis goes back to ancient Epicureanism, the second receives its initial modern expression in Richard Cumberland's 1672 *A Treatise of the Laws of Nature*. More distinctively, however, and in contrast to secular versions of utilitarianism, Anglican utilitarians contend that morality needs God, particularly for a satisfactory account of moral obligation and for a solution to the problem of conflict between private and public happiness.

The analysis of Anglican utilitarianism offered here strives to make it comprehensible

by seeing it as the synthesis of two currents of thought, both of which developed in the seventeenth century: Protestant natural law theory and the modern revival of Epicureanism. The preceding intellectual labors expressed in these traditions enabled the Anglican utilitarians to articulate a novel ethical system, one that had an influential life in Britain. Protestant natural law theory and modern Epicureanism are examined through analyses of key late-seventeenth-century figures: Cumberland and, the most important influence on the development of Anglican utilitarianism, John Locke. The section then moves on to interpret the works of George Berkeley, Francis Hutcheson, David Hume, and the Anglican utilitarians vis-à-vis natural law, Epicureanism, and the establishment of utilitarian ideas and arguments.

Cumberland and Locke

Cumberland

It would be too weak to say that the most famous of the classical utilitarians – Jeremy Bentham and John Stuart Mill – *rejected* the idea of a law of nature. Rather, they *ridiculed* it. Bentham asserted that the "pretended *law of nature*" was nothing but "an obscure phantom."[4] John Stuart Mill decried the "imaginary law of the imaginary being Nature."[5] Instead of seeing law as the fundamental organizing principle of morality, the classical utilitarians took law to be derivative of a more fundamental notion: the good.

Perhaps surprisingly, then, when we examine the history of utilitarianism in Britain prior to Bentham, we discover that utilitarian theories were natural law theories, in which God is the legislator. Richard Cumberland (1631–1718) can be seen as the first to put a utilitarian view in natural law garb.

Natural law morality emphasizes that morality is a universal law, imposed by a law-giver (typically God), and knowable by reason alone without the aid of revelation (thus "natural"). Though the idea that there is a universal moral law extends back to the Stoics, natural law morality received a seminal development at the hands of Thomas Aquinas (and still remains an important part of Catholic moral philosophy today). In the period after the Protestant Reformation, however, a new school of natural law morality, Protestant natural law, began.[6]

One landmark in Protestant natural law theory was the publication of Cumberland's *A Treatise of the Laws of Nature*. For Cumberland, the demands of God's law (i.e., morality) are reducible to one: promote the *common good* of rational agents, namely, the honor of God and the happiness of humans.

On Cumberland's account, we can know that the common good is an obligatory end independently of scriptural revelation. This knowledge arises through careful observation of nature, because the will of God is "naturally known" in these matters most clearly

from its effects.[7] That is, we can infer that God has willed us to pursue the common good because, for the agent, "[i]nnumerable evils . . . *naturally attend* every Action *injurious* to others" and various goods accompany every action beneficial to the common good.[8]

Once we recognize that God has willed us to promote the common good, his authority as the "Governour of the World" makes this a *law* for us.[9] God's will, in other words, makes the difference between its being merely *reasonable* to pursue the common good and its being something that we are *bound* to promote. Without God, obligation – and therefore morality – is impossible.

What is the significance of Cumberland's account of the common good, particularly for the history of utilitarianism? Two things seem noteworthy. First, it was *not* unusual to claim that the end of natural law is the common good (indeed, Aquinas does so). However, Cumberland makes the novel claim that the common good is something that is a *sum* of the good of the individuals that make up the community of rational beings – it provides a basis for the comparison of the moral worth of two acts.[10] In addition, Cumberland suggests that the criterion of right action is how much an action's *consequences* promote the common good (though he is not consistent on this point – it gets a much clearer expression in the Anglican utilitarians).[11] These positions open conceptual space that utilitarian theories can occupy.

Second, the common good provides a *public standard* that makes shared reasoning about right and wrong possible. This serves as a reaction against an Aristotelian reliance on the judgment of the practically wise person as the final determiner of right or wrong in particular instances. Cumberland called this feature of the common good its "greatest Advantage of all." From

> the very Nature of the common Good . . . a certain Rule or Measure is afforded to the prudent Man's Judgment, by the help whereof he may ascertain that just Measure in his Actions and Affections, in which Virtue consists. This Task Aristotle has assign'd to the Judgment of the Prudent, in his Definition of Virtue, but has not pointed out the Rule by which such Judgment is to be form'd.[12]

Cumberland, following Hugo Grotius, claims that rules take priority over judgment, rather than claiming, as Aristotle does, that moral rules are imperfect measures of the judgment of the practically wise person. This theme takes a new form in Bentham's subsequent call for an "external standard" or "extrinsic ground" for moral judgment and his complaints about the principle of sympathy and antipathy – "that principle which approves or disapproves of certain actions . . . merely because a man finds himself disposed to approve or disapprove of them."[13]

Locke

It might appear a bit odd that Locke (1632–1704) should figure so prominently in the story of utilitarianism since he denies that right action is determined by appeal to a utilitarian principle such as the principle that right acts are those whose consequences promote the common good. This denial rests on a different determination of what nature reveals that God wills for his creatures – what law God establishes for humans. In Locke's account, the most fundamental law of nature is the "*Peace and Preservation of all Mankind*," which he then divides into two: that each individual is "*bound to preserve himself*" and, when "his own Preservation comes not in competition," to "*preserve the rest of Mankind*."[14] Further duties and rights are those necessary to the fulfillment of these laws, so the right to property, for instance, makes preservation of oneself – and thus the fulfillment of God's will for us – possible. It is *not* the case that the justifications for respecting property rights or prohibiting suicide depend on an appeal to the common good. Locke does not take over Cumberland's innovation in natural law, yet the distance is perhaps not so great between the thoughts that God wills the preservation of humanity and that God wills the happiness of humanity.

Though Locke disagrees with Cumberland and Anglican utilitarians about the proximate criteria for right action, he nevertheless establishes and, with the growth of his authority in England, legitimates the philosophical path that the utilitarians will largely follow. In particular, he does two important things. He defends a voluntarist theory of obligation to natural law in which God plays an essential role and he argues for various egoistic and hedonistic theses. I briefly discuss each in turn.

Locke's theory of moral obligation was taken over by Anglican utilitarians, among others. Locke wants to claim both that our obligation to the moral law ultimately rests on our obligation to obey God's will (the source of that law) and, against Hobbes, that our obligation to obey God's will (and the natural law that God wills) does not rest simply on God's irresistible power.[15] While God's rewards and punishments motivate us to obey God's law, obligation to God's law has its source in God's authority over us. If one pushes further, and asks on what that authority is founded, Locke emphasizes that God is our maker and that the "making relationship" carries with it obligations of the made to the maker and rights of the maker in the made.[16]

Locke is at his most important for the history of utilitarianism in his defense (and legitimation) of various egoistic and hedonistic theses. As noted above, one key tenet of utilitarianism – all things are valuable only insofar as they promote happiness (i.e., pleasure) – has a very long history, stretching back to Epicurus and the Hellenistic philosophical school that grew from his teachings.

It seems that Locke was spurred into adopting an egoistic and hedonistic theory through the influence of a variety of French neo-Epicureans, most prominently Pierre Gassendi.[17] Locke's version of egoism and hedonism includes three interrelated theses.

First, a thesis about value, namely that we call 'good' what has an aptness to produce pleasure and 'evil' what has an aptness to produce pain.[18] In one of his notebook entries from 1676, Locke claims that those things are good which make up our happiness (i.e., pleasure) while those things that procure pleasure, but are not themselves pleasant, are good only in a secondary sense. Under the latter, Locke includes the useful (*utile*) and virtue or the honorable (*honestum*). He argues that goods like money and temperance, "were they not ordained by God to procure the *jucundum* [pleasure] and be a means to help us to happiness . . . I do not see how they would be reckoned good at all."[19] Similarly, gluttony would not be counted a vice if it did not produce pain.

Second, Locke's treatment of happiness, unlike Cumberland's, clearly presents happiness as reducible to pleasure and the absence of pain. Happiness is a reckoning of pleasures and pains. Third, an egoist thesis about motivation: everyone pursues individual happiness – we are exclusively motivated to act by our desire to get pleasure and avoid pain.[20]

One important modern innovation in Epicurean hedonism, however, was its Christianization. In hedonistic theories, the main deliberative problem is often trying to determine what things promote and what things reduce pleasure. Locke, Gassendi, and others argued for the essential roles that the afterlife and God's sanctions in it play in organizing our calculations of happiness. Our true happiness does not lie in the immediate pleasures of this life, but in following God's will so that we garner eternal happiness and avoid eternal pain.[21] Indeed, as Locke emphasizes, without God, human beings would find living together exceptionally difficult. If our desire for pleasure were left unchecked, it would undermine morality and social life.[22]

The "aim and tendency" of Locke's philosophy, according to Edmund Law, "is no other than to reduce the foundations of our Knowledge, and our Happiness, to that original *Simplicity* which Nature seems to observe in all her Works."[23] The Lockean emphasis on metaphysical simplicity, naturalism, and epistemic humility would shape much subsequent discussion in moral philosophy – both as a source of inspiration and as an object for attack.

Eighteenth-century developments

Berkeley

In his three discourses on "Passive Obedience," delivered to students in 1711 (or possibly early 1712) at Trinity College, Dublin, George Berkeley (later Bishop Berkeley) (1685–1753) makes a case for the absolute duty of submitting to the supreme civil power.[24] In so doing, he allies himself with Tories and more conservative Anglicans squarely against, among many others, Locke and his Whig colleagues, who argued for

the right to resistance of sovereign authority. More controversially, and more damaging to Berkeley's career, the doctrine of passive obedience was frequently identified with "Jacobites" (i.e., supporters of the deposed James II and opponents of the Glorious Revolution), even though there are good reasons for believing that Berkeley was no Jacobite. In making the argument for passive obedience, Berkeley relies on natural law to formulate a clear utilitarian position. Berkeley's position is not representative of how utilitarian thought (in its theological or secular varieties) would ultimately develop in Britain; thus, though he is an Anglican and a utilitarian, I separate him from the mainstream Anglican utilitarianism centered in Cambridge. Nevertheless, his theory is important both for its originality and because it provides a useful historical instance showing that utilitarian ideas can be employed for conservative (indeed, potentially reactionary) ends.

The core of Berkeley's argument – much of which is consistent with Locke's general views – goes as follows. It is natural for human beings to regard things in their relation to self-love, naming them 'good' if they promote one's own happiness and 'evil' if they reduce it. As we come to maturity, we are able to comprehend that immediate pleasures often lead to pain and vice versa. Every person thus recognizes that every "reasonable man" ought to act in the way that most contributes to "his eternal interest."[25] Since we know from natural reason that there is a "sovereign, omniscient Spirit, who alone can make us for ever happy, or for ever miserable: it plainly follows that a conformity to his will, and not any prospect of temporal advantage, is the sole rule whereby every man who acts up to the principles of reason must govern and square his actions."[26]

Conforming to God's will means following the laws of nature that issue from God's will. But in order to determine by natural reason (rather than by revelation) the laws of nature, one needs to determine first what God intended in making the laws, that is, what "that end is, which he designs should be carried on by human actions."[27] Berkeley, like Cumberland, argues that God intends the "good of men" and that since no one man qua man is entitled to more than another, God intends the "general well-being of all men, of all nations, of all ages of the world."[28] Unlike Cumberland, the bulk of whose argument attempts to reveal God's intentions empirically by detailing the sanctions associated with specific kinds of actions, Berkeley relies on an *a priori* argument from God's infinite goodness: "as God is a being of infinite goodness, it is plain the end he proposes is good. But God enjoying in himself all possible perfection, it follows that it is not his own good, but that of his creatures." After discovering God's end, we are able to identify the means that "necessarily promote" achieving this end, namely, observance of the laws of nature.[29] Anglican utilitarians make arguments about God's intentions more similar to Berkeley's than to Cumberland's.

Another notable feature of Berkeley's account is what we would call his "rule utilitarianism."[30] Some of the laws of nature are absolute, that is, they must be followed in all cases, without exception. All the absolute laws are negative precepts and include

commands not to commit adultery, murder, steal, lie, or resist the supreme power (i.e., the law of passive obedience). Though these are laws of nature because they are means to promoting "the public weal," it is *not* the case that we are permitted to overlook the laws in particular cases in order to refer directly to the "good of men."[31] The principal reason for this restriction is that without determinate laws of nature to guide us, everyone is left to his or her own judgment to a degree that Berkeley thinks is untenable and leads to chaos. The general good is simply too indeterminate an end to curtail sufficiently the bad tendencies of creatures like us. In addition, we are unable to calculate the consequences of actions for the general good, and even if we could, we would lack the time to do it. This rigorism about the moral law serves the purpose of defending passive obedience against those who would claim that some circumstances or acts by the sovereign could justify rebellion. In making that case, Berkeley anticipates some of the arguments that would be made in subsequent centuries against act utilitarianism.[32]

Scottish developments

Utilitarianism did not take hold in Scotland the way it did in England. Nevertheless, two Scottish philosophers, Francis Hutcheson and David Hume, provided some conceptual support for the later development of secular forms of utilitarianism, though neither should be considered, as is sometimes done, a utilitarian.

Hutcheson (1694–1746) was a Scottish-Irish Presbyterian who profoundly shaped the Scottish Enlightenment through his writings and his teaching as professor of moral philosophy at the University of Glasgow. Hutcheson has sometimes been taken to be a utilitarian thinker because of the importance of the "general Good" in his moral and political thought.[33] So, for instance, we say that someone has a right to do, possess, or demand something when it "would in the whole tend to the general Good."[34] We call God morally good "when we apprehend that his whole Providence tends to the universal Happiness of his Creatures."[35] Moreover, Hutcheson coined the utilitarian phrase "the greatest Happiness for the greatest Numbers" when talking about a criterion to use in choosing among different possible actions.[36]

The suggestion that Hutcheson is a utilitarian, however, misses some crucial differences. For Hutcheson, unlike for the more standard utilitarians of the period, what is morally important is not happiness. Actions are ultimately evaluated not by their consequences for happiness, but by the underlying motives that produced the action. Hutcheson contends that we approve of agents' actions (including God's) when they are done from the motive of benevolence – all virtue (even justice) is an expression of benevolence. Both utilitarians and Hutcheson emphasize the public good, but for Hutcheson, this is because an action's promoting the public good provides evidence for the agent's benevolence in acting, not because public happiness is intrinsically good. Hutcheson finds the source of moral goodness in the character of *agents*, not in

objectively desirable *states of affairs* (i.e., happiness).

Though rejecting egoistic hedonism and inspired by many aspects of Hutcheson's thought, Hume (1711–1776) remained more fully Epicurean.[37] Hume's development of Epicurean themes, particularly his well-known theory of utility as the source of our valuing justice and his rejection of key tenets of natural law and natural religion, made him an important forerunner of Benthamite utilitarianism. Nevertheless, there are a number of reasons not to consider Hume a utilitarian. First, like Hutcheson, Hume argues that we judge actions by the motives they express, not by the consequences they produce. Second, it is not clear that by 'utility' he means pleasure that can be summed or aggregated rather than something like social order or stability. Third, he claims some virtues (e.g., cheerfulness, politeness) are valued because of their immediate agreeability rather than their public utility. Finally, it is not obvious that Hume defends any normative position in morality at all, let alone a monistic one like the principle of utility; he appears more focused on describing moral phenomena as part of a "science of man." Hume's legacy for utilitarianism is a complicated one, as Bentham's ambivalence toward Hume indicates.[38]

Anglican utilitarians

By the 1730s, mainstream Anglican theology stressed conduct (rather than belief in doctrine) as the foundation of the good Christian life and knowledge derived from natural reason (rather than from revelation). These two emphases led to increased interest in moral philosophy, and Anglican utilitarianism attracted a dominant share of that interest. It was the only instance in the eighteenth century of utilitarianism gaining prominence among an institutionally established elite. The books of Anglican utilitarianism were written mostly by Cambridge men and were used to teach many succeeding generations, who often went on to influential careers in the Church and in British political life.

Evidence of this preeminence comes from William Whewell – philosopher, antagonist of Mill's, master of Cambridge's Trinity College – who noted in 1852 that the utilitarian philosophy of Gay, Tucker, and Paley is "*the* scheme of morality which has been taught in this University for the last century."[39] Contrary to any historically naive expectations that Bentham's secular utilitarianism "replaced" religious utilitarianism, the importance of Anglican utilitarianism became even greater *after* the 1789 publication of Bentham's *Introduction to the Principles of Morals and Legislation* due to the British reaction against the natural rights theories that were associated with political radicalism and the revolution in France. In particular, Paley's *The Principles of Moral and Political Philosophy* (1785), which was part of the curriculum at Cambridge well into the nineteenth century, garnered the compliments of institutional endorsement and critical attack and remained far more influential than Bentham's work for a number of decades.

Anglican utilitarianism is both continuous with and in opposition to the secular

utilitarianism of Bentham and the Mills (John Stuart Mill and his father James). There are continuities in the criterion of morality, moral psychology, and value theory, but distinct differences in moral obligation and in how revisionary moral philosophy was taken to be. Greater familiarity with Anglican utilitarianism offers, among other things, a richer appreciation of the relations of Benthamite utilitarianism to mainstream eighteenth-century British moral and political thought.

Though there are differences in emphasis, sophistication, and (in a few instances) doctrine among those I have named "Anglican utilitarians," the key themes of Anglican utilitarianism all get expressed in John Gay's brief "Preliminary Dissertation Concerning the Fundamental Principle of Virtue or Morality" and it thereby provides a convenient focus for this chapter (even if Paley's work remains the most eloquent and influential presentation of these themes). The work was originally published anonymously as a preface to Edmund Law's 1731 translation of William King's *Essay on the Origin of Evil*. It is well known for being the first attempt to synthesize two things that would become particularly important in nineteenth-century thought: utilitarianism and associationism (i.e., the Lockean attempt to explain mental phenomena by showing them to be the ultimate product of simpler components of experience connected to each other through associations of, e.g., resemblance and contiguity). The work's relative impact far outstrips its length.

Gay (1699–1745) was educated at Cambridge, where he was a fellow and tutor from 1724 to 1730. Law's judgment that Gay's knowledge of the Bible and of the works of Locke was unrivaled expressed the high opinion Gay's colleagues at Cambridge had of him (and the veneration they had for Locke).[40]

There are three features of Anglican utilitarianism deserving special notice that get expressed in Gay's work. First, the Anglican utilitarians develop (in some novel ways) Epicurean hedonism. Second, they defend a utilitarian criterion of right action. Third, they argue that moral obligation requires God and God's sanctions. We will address each in turn.

Following Locke and other modern Epicureans, Gay contends that "Happiness, private Happiness, is the proper or ultimate End of all our Actions whatever."[41] All action originates in pursuit of pleasure or avoidance of pain. And, as Gay's successors argued, not only are we motivated by pleasure and pain, pleasure and pain act as the ultimate source of moral value.

This idea and its implications get developed and spelled out more clearly by Gay's successors. Law thought that because there are no "original pleasures or pains beside sensitive ones . . . any innate intellectual determination, or Moral Principle wholly underived from and naturally independent of these, seems an *impossibility*."[42] Soame Jenyns had little but contempt for moral principles "wholly underived" from pleasure and pain: "They who extol the truth, beauty, and harmony of Virtue, exclusive of its

consequences, deal but in pompous nonsense."[43] Jenyns's ridicule expresses the shared Lockean (and Epicurean) attitude that Anglican utilitarians took toward Lord Shaftesbury and toward ethical rationalists like Samuel Clarke, William Wollaston, and John Balguy who understood an act's rightness in terms of the relations of ideas expressed in that action.[44] Anglican utilitarians argued that there is nothing intrinsically valuable in truth, beauty, or virtue (or in acts insofar as they possess those qualities).

Gay complicates Epicurean egoism and hedonism, however, by agreeing with Hutcheson and others that we *do* approve of and pursue things like virtue or money for their own sakes. He nevertheless reconciles these phenomena with his hedonism through associationism. We originally love things like money as a means to happiness. Eventually, however, the association of money with pleasure can become so strong that we begin to pursue money for its own sake, not merely as a means to something else – the money itself becomes pleasurable. The same thing often happens with virtue – we begin by loving it for the personal pleasure it brings, and end up, through the power of association, loving it because it itself is pleasurable. So, the power of association explains how we move from seeing virtue as a means to happiness to seeing it as itself pleasurable, to be pursued for its own sake. Virtue is not (and should not be) valued independently of its connection to pleasure, but it can be valued for its own sake *in the sense* of being an ultimate rather than a subordinate end for action.

Gay's short dissertation also addresses the themes of a criterion for right action (or "criterion of virtue") and moral obligation. These are tightly related for him, as they are for his fellow utilitarians.

Gay defines virtue as "*the Conformity to a Rule of Life, directing the Actions of all rational Creatures with respect to each other's Happiness; to which Conformity every one in all Cases is obliged: and every one that does so conform, is or ought to be approved of, esteemed, and loved for so doing*."[45] Virtue, in other words, is conformity to an obligatory rule of life. The criterion of virtue is "what it is which denominates any Action virtuous," and to inquire after it is "to enquire what that Rule of Life is to which we are oblig'd to conform" (one clearly sees here the influences of the natural law tradition).[46]

To know what counts as virtue is, then, to know what rule of life is obligatory for us. Gay defines obligation as "*the necessity of doing or omitting any Action in order to be happy*: i.e., when there is such a relation between an Agent and an Action that the Agent cannot be happy without doing or omitting that Action, then the Agent is said to be *obliged* to do or omit that Action."[47] Gay claims that a "full and complete Obligation . . . can only be that arising from the Authority of God; because God only can in all Cases make a Man happy or miserable: and therefore, since we are always obliged to that conformity call'd Virtue, it is evident that the immediate Rule or Criterion of it is the Will of God."[48] So, on Gay's view, the criterion of virtue is something that we must be

obliged to follow in *all* cases. Given Gay's definition of obligation, in which the motive of self-interest is what obligates us, and given that only God can in *all* cases make a man happy or miserable, only God can oblige in all cases (though Gay notes the contributions that other sanctions – natural, virtuous, civil – make to obligation).[49] Thus we must be obliged to do whatever God wills us to do.

The next natural question: What does God will for us? Like Berkeley, Gay avoids the laborious attempt by Cumberland to identify God's will for us through detailed empirical study of the world (particularly the sanctions associated with actions). Instead, he argues from God's nature – God's perfect happiness and his goodness – that it is "evident" God could have "no other Design in creating Mankind than *their* Happiness."[50] Gay goes on to claim that God must also will the means to human happiness. Thus, my own conduct "as far as it may be a means of the Happiness of Mankind, should be such."[51] For Gay, then, the ultimate test of actions – the criterion of virtue – is the will of God. The more proximate test of the morality of actions is that which God wills, namely, the happiness of humanity.

It is worth reflecting for a moment on the idea – central to Anglican utilitarianism – that God wills our happiness: Why should we believe that God's principal goal in creation is the happiness of his creatures? Why shouldn't we think that this position overemphasizes God's benevolence (and does not acknowledge other features of God, like his vengefulness)? Or that God's ends are impenetrable to us – that God is mysterious? Or that God has lots of ends in designing humanity, only one of which is humanity's happiness? It is interesting how little effort seems to go into resolving these problems (see, for instance, Paley's rather cursory arguments in *The Principles of Moral and Political Philosophy*).[52] This may indicate a general agreement concerning God's ends and our determination of those ends among readers of the Anglican utilitarians.

Leslie Stephen's quip that Bentham is "Paley *minus* a belief in hell-fire" captures the major discontinuity between secular utilitarianism and the theory defended by Gay, Paley, and the other Anglican utilitarians: God.[53] God plays two vital (and closely related) roles in Anglican utilitarianism. First, God provides reasons for favoring public interest when it conflicts with private interest. Second, on the assumption that promoting the public interest is the most basic demand of morality, God thereby provides a ready answer to the question: "Why should I be moral?"

In order to show God's necessity for choosing public over private interest, Gay presents a dilemma to an imagined secular utilitarian theory that omits God and makes the good of humanity the ultimate criterion of virtue. Proponents of such a theory "must either allow that Virtue [i.e., the public interest] is not in all Cases obligatory (contrary to the Idea which all or most Men have of it) or they must say that the Good of Mankind is a sufficient Obligation."[54] Since it seems obvious to Gay that virtue must obligate, he focuses on the second horn of this dilemma. In principle, the public good might obligate

in one of two ways: either it is itself capable of generating obligation even when it is in conflict with private happiness or there is never a conflict between self-interest and the "Good of Mankind." But Gay rejects both possibilities.

Gay rejects the first possibility because the public good, unlike private happiness, does not provide a reason for action that is sufficient to obligate. As Gay puts it, "But how can the Good of Mankind be any Obligation to me, when perhaps in particular Cases, such as laying down my Life, or the like, it is contrary to my Happiness[?]"[55] (The presupposition that our obligations must align with our happiness was, unsurprisingly, challenged by Gay's contemporaries.)[56] Gay rejects the second possibility – that the public good is always in one's self-interest even in the absence of divine sanctions – because there are clear cases in which acting for the public good would mean sacrificing one's private happiness if it were not for the rewards of heaven. The conclusion of Gay's dilemma for his imagined secular utilitarian theorist is that, without God, there is no coherent way to make the good of humanity the ultimate criterion of virtue.

So, without God, in cases of conflict between public and private interest, one has no obligation to be moral. Indeed, the atheist would seem to have good reason to pursue his pleasure. But the inclusion of God solves this problem. If it is assumed that God generously rewards people who promote the public good and severely punishes those who frustrate it, then it is guaranteed that promoting the public good will be in one's long-term (if not one's earthly) self-interest. And on the assumption that promoting the public interest is required by morality, then by the same token God provides an answer to the question: "Why should I be moral?"[57]

In providing a solution to the problem of moral conflict, God brings order to the world – a world that without God would be disordered. Gay's skepticism about secular utilitarianism is grounded on skepticism that one can offer an egoist (e.g., a typical human being) reasons to be moral.

We can say, in summary, that utilitarianism attracted this group of thinkers for a number of reasons. Among them are utilitarianism's metaphysical and epistemic simplicity, its compatibility with religious rationalism (e.g., it can be articulated without reference to contentious appeal to scripture or to controversial theological claims), and its capacity to offer a defense of the social and political status quo (e.g., the legitimacy of established authorities).

French utilitarianism

As in seventeenth- and eighteenth-century England, so too in France: A variety of egoist and hedonist theses were widely and influentially defended. And these helped generate, in at least a few cases, utilitarian moral theories.

An essential contrast between eighteenth-century English and French egoists, as

described by J. B. Schneewind, "was their understanding of the universe in which self-love should direct us." While the English theorists saw egoism as operating in "a divinely ordered universe," in which morality is expressed in universal natural law, willed by God, the French egoists such as Claude Helvétius and Baron d'Holbach rejected that.[58] The refusal to invoke God or God's providence in utilitarian morality gave these theories a distinctive character that would play an essential role in the development of Benthamite utilitarianism. Here, we will focus on the most important of those thinkers: Helvétius (1715–1771).

Henry Sidgwick, perhaps the greatest of the "classical utilitarians," contends that "the premises of Bentham are all clearly given by Helvétius."[59] Helvétius was part of the radical French Enlightenment that included Denis Diderot, Voltaire, and d'Holbach. His most important work, *De l'esprit*, was published in 1758, publicly burned, and the source of much controversy (so much so that it has been argued that it was more important in building the intellectual scaffolding of the French Revolution than was Rousseau's *Social Contract*).[60]

His account of human nature draws especially from Locke (whose influence in France at this time was profound) and modern Epicureanism.[61] Among other things, he emphasizes the continuity of human beings with animals and denies innate evil or wickedness, both claims that put him in opposition to a variety of religious anthropologies. Helvétius also takes Locke's anti-innatism in radical directions, arguing for the great power of circumstance and experience in the shaping of human beings. This led him to emphasize, as we will see, the centrality of legislation and education for realizing the ends of morality. It was also part of a radical egalitarianism that attributed most differences between humans to differences in their situations rather than in their innate capacities.[62]

Helvétius clearly defends a utilitarian criterion of morality and politics. Public utility "is the principle on which all human virtues are founded."[63] So, for instance, a man is just "when all his actions tend to the public welfare."[64] It is the standard that should determine what counts as virtue and vice. It is, moreover, the standard that typically underlies our approval of virtues, even if we claim that it does not.[65]

Helvétius combines this commitment to a utilitarian criterion of virtue with an Epicurean egoism and hedonism. Such egoism and hedonism involve two familiar claims. First, he asserts that self-love or personal interest (i.e., pleasure and freedom from pain) is the "principle of all our actions." As he put the point: "[God] seems also to have said to man . . . I place thee under the guardianship of pleasure and pain: both shall watch over thy thoughts, and thy actions; they shall beget thy passions, excite thy friendship, thy tenderness, thine aversion, thy rage; they shall kindle thy desires, thy fears, thy hopes . . ."[66] Second, he argues that all value can be understood as grounded in pleasure and pain. There is no standard for good and evil other than our responses to it, and our moral

approval and disapproval – our judgment of good and bad, right and wrong – is determined by the pleasure and pain objects cause.

Yet, unlike the Christian Epicureanism that Locke and others espouse, Helvétius does not wish to invoke God and God's sanctions in the afterlife in order to explain how potential conflicts of personal and public interest get resolved. This leaves him with the challenge that Gay articulates for a secular utilitarian theory: Without God, what reasons are there for why an individual should sacrifice her interest for the sake of general utility or virtue?

Here, Helvétius makes a move that would be decisive for the future of utilitarianism, particularly in its establishing a conceptual structure for Bentham: "I say, that all men tend only towards their happiness; that it is a tendency from which they cannot be diverted . . . consequently, it is only by incorporating personal and general interest, that they can be rendered virtuous. This being granted, morality is evidently no more than a frivolous science, unless blended with policy and legislation."[67] As indicated at the end of this quote, Helvétius turned from God to human legislation and education in order to explain how individual and general interests could be made to cohere. It is through sound policy and legislation (not through God's providence) that our pursuit of individual interest most consistently promotes the general interest, and it is bad policy and legislation that exacerbates conflict between individual and general interest.

Helvétius notes that moralists and divines have made the mistake of attributing the bad actions of human beings to their natures rather than to their circumstances – circumstances frequently created by legislators. A nation's legislation is "the root whence its vices arise." Helvétius admonishes those who continually emphasize the "malignity of mankind" and notes that human beings "are not cruel and perfidious, but carried away by their own interest." Rather than complain about the wickedness of humanity, moralists ought to decry "the ignorance of legislators, who have always placed private interest in opposition to the general interest."[68]

Helvétius calls for political reform as a vehicle for moral reform. By changing the circumstances in which self-interested agents act, one can increase the likelihood that acting for the general utility will be in the interest of the individual agent. Particularly effective in this regard is making pain accompany vicious acts and pleasure accompany virtuous ones. As he puts it, "the love of pleasure . . . is a bridle by which the passions of individuals might always be directed to the public good," and "true virtue is founded on the love of esteem and glory, and the fear of contempt, which is more terrible than death itself."[69]

Helvétius presents both a thoroughly naturalistic utilitarian theory and a model for utilitarianism as an instrument of reform (this reforming brand of utilitarian thought gets further influential expression in the work of the Italian legal reformer, Cesare Beccaria, whose book *Of Crimes and Punishments* (*Dei delitti e delle pene*, 1764) defended a

utilitarian account of law that was important for Bentham, among others). Helvétius employed a utilitarian criterion of right action as a measure for the moral and political worth of prevailing institutions and practices. This was one tradition that Bentham drew upon in developing his own ideas about reforming law, politics, and morality.

Conclusion

The differences between Anglican and secular utilitarian theories might justify two separate histories rather than the single one offered here, particularly when one thinks about the ways in which different versions of utilitarianism engaged with and developed from moral, political, and legal problems and concerns. So, for instance, one of the reasons to use 'Anglican' rather than 'theological' (as has been done in the past) to describe one brand of utilitarianism is because it brings out how deeply connected this moral philosophy was to the *institutional* life of church and university. This makes unsurprising Anglican utilitarianism's roles in defending moderate conceptions of what it meant to be a Christian, of the centrality of reason in religion, and of the legitimacy of the political and social status quo. In this guise, utilitarian thought did little to advocate for fundamental reform – rather it tried to make existing practices more intelligible and demonstrate why they were desirable. Secular versions of utilitarianism, alternatively, brought out some of the more radical possibilities latent in utilitarian thought – such as the idea that every person's (or sentient being's) happiness deserved equal consideration to that of every other person's. It thereby largely denied the prima facie justification that history and tradition offered to existing practices and institutions.

In an important sense, however, both Anglican and secular utilitarians share important attitudes about morality – a basic (Epicurean) intuition that the experiences of pleasure and pain are the fundamental ones for human beings, desire for a public or "external" standard for moral reasoning, suspicion about moral entities whose reality is difficult to ascertain (e.g., "the truth, beauty, and harmony of Virtue"), and a Lockean temperament favoring simple explanation and epistemic humility. These shared attitudes, combined with shared commitment to a variety of philosophical theses and arguments, support the validity of a shared history.

Thanks to Ben Eggleston and Dale Miller for their excellent editorial supervision and to Jim Crimmins and James Harris for helpful comments on an earlier draft of this chapter.

Notes

1. Thanks to Jim Crimmins for discussion of this point.

2. See the conclusion for reasons to prefer the term 'Anglican utilitarianism' over the commonly used expression 'theological utilitarianism'.

3. The best collection of their works (with helpful introductory essays) is Crimmins, *Utilitarians and Religion*.

4. Bentham, *IPML*, p. 298, n. a2.

5. J. S. Mill, *Auguste Comte and Positivism*, *Collected Works*, vol. X, p. 299.

6. For overviews of Protestant natural law, see, e.g., Haakonssen, *Natural Law and Moral Philosophy*; and Tuck, *Natural Rights Theories*.

7. Cumberland, *A Treatise of the Laws of Nature*, p. 536. Cumberland also offers, somewhat half-heartedly, a few *a priori* arguments for knowing the content of God's will; see pp. 537–541.

8. Cumberland, *A Treatise of the Laws of Nature*, p. 333.

9. Cumberland, *A Treatise of the Laws of Nature*, p. 536.

10. Cumberland, *A Treatise of the Laws of Nature*, pp. 355–356, p. 537, and p. 575.

11. Cumberland, *A Treatise of the Laws of Nature*, pp. 505–508. For discussion of Cumberland on these and other points, see Schneewind, *The Invention of Autonomy*, chapter 6.

12. Cumberland, *A Treatise of the Laws of Nature*, p. 275.

13. Bentham, *IPML*, chapter 2, sections 14 and 11 (both on p. 25).

14. Locke, *Two Treatises of Government*, p. 271 (in book II, chapter 2, section 6).

15. For discussion see Darwall, "Norm and Normativity," p. 991.

16. For detailed discussion, see Tully, *A Discourse on Property*, pp. 38–42. For problems with Locke's approach to obligation, see Darwall, "Norm and Normativity."

17. Driscoll, "The Influence of Gassendi on Locke's Hedonism." It should also be noted that various versions of early modern Augustinianism (e.g., that of La Rochefoucauld) also presented human beings as egoistic and hedonistic – a result of the Fall. See Force, *Self-Interest before Adam Smith*, for further discussion of the relations of Epicureanism and Augustinianism in this period.

18. Locke, *ECHU*, p. 259 (in book II, chapter 21, section 42).

19. Locke, *Political Essays*, p. 241. See also Locke, *ECHU*, p. 229 (in book II, chapter 20, section 2) and p. 259 (in book II, chapter 21, section 42).

20. It should be noted that Locke appears to change his position in the second edition of *ECHU*, p. 249 (in book II, chapter 21, section 29), emphasizing "uneasiness" as the motive for action (perhaps influenced by Malebranche) rather than the pursuit of pleasure and the avoidance of pain. I thank James Harris for pointing to this shift in Locke's views.

21. With Gassendi, and in contrast to Hobbes, it is a bad afterlife, not death, that presents us with the greatest of all evils – see Driscoll, "The Influence of Gassendi on Locke's Hedonism," pp. 106–107.

22. Locke, *ECHU*, p. 75 (in book I, chapter 3, section 13).

23. Law, "The Nature and Obligations of Man," p. lx.

24. For discussion of the institutional and political contexts of Berkeley's discourses, see Berman, "The Jacobitism of Berkeley's Passive Obedience," and I. C. Ross, "Was Berkeley a Jacobite?"

25. Berkeley, *Passive Obedience*, p. 5.

26. Berkeley, *Passive Obedience*, pp. 7–8.

27. Berkeley, *Passive Obedience*, p. 6.

28. Berkeley, *Passive Obedience*, p. 6.

29. Berkeley, *Passive Obedience*, p. 6 and p. 18.

30. For discussion, see Dale Miller in this volume (Chapter 7).

31. Berkeley, *Passive Obedience*, p. 16.

32. For discussion, see Ben Eggleston in this volume (Chapter 6).

33. Hutcheson, *An Inquiry into the Original*, p. 182.

34. Hutcheson, *An Inquiry into the Original*, p. 182.

35. Hutcheson, *An Inquiry into the Original*, p. 181.

36. Hutcheson, *An Inquiry into the Original*, p. 125.

37. For Hutcheson's connections to Hume, see especially Norton, *David Hume*. For Hume's Epicureanism, see J. Moore, "The Eclectic Stoic, the Mitigated Sceptic," and Rosen, *Classical Utilitarianism from Hume to Mill*. For a more skeptical account of Hume's Epicureanism, see James Harris, "The Epicurean in Hume."

38. For a very helpful discussion of Hume's place in utilitarianism and of Bentham's ambivalence toward him, see Crimmins, *Utilitarian Philosophy and Politics*, chapter 3.

39. Whewell, *Lectures on the History of Moral Philosophy*, p. 137; italics added.

40. Jonathan Harris, "John Gay."

41. Gay, "Preliminary Dissertation Concerning the Fundamental Principle of Virtue or Morality," p. lxvi.

42. Law, "The Nature and Obligations of Man," p. lix.

43. Jenyns, *A Free Inquiry into the Nature and Origin of Evil*, p. 123. See also J. Brown, *Essays on the Characteristics*, pp. 64–65.

44. E.g.: "To give Pain, without cause, to a sensible Creature, is an Action self-evidently wrong; as being directly repugnant to the Nature of the Object, and the Circumstances of the Agent" (Balguy, *The Foundation of Moral Goodness*, p. 38).

45. Gay, "Preliminary Dissertation Concerning the Fundamental Principle of Virtue or Morality," p. xxxvi.

46. Gay, "Preliminary Dissertation Concerning the Fundamental Principle of Virtue or Morality," p. xxviii and p. xxxvii.

47. Gay, "Preliminary Dissertation Concerning the Fundamental Principle of Virtue or Morality," p. xxxvii.

48. Gay, "Preliminary Dissertation Concerning the Fundamental Principle of Virtue or Morality," p. xxxviii.

49. For discussion, see Darwall, "Norm and Normativity," pp. 993–994.

50. Gay, "Preliminary Dissertation Concerning the Fundamental Principle of Virtue or Morality," pp. xxxviii–xxxix.

51. Gay, "Preliminary Dissertation Concerning the Fundamental Principle of Virtue or Morality," p. xxxix.

52. Paley, *The Principles of Moral and Political Philosophy*, book II, chapter 5 ("The Divine Benevolence") (pp. 39–42).

53. L. Stephen, *History of English Thought in the Eighteenth Century*, p. 125. Thanks

to Ben Eggleston for very helpful suggestions for this discussion.

54. Gay, "Preliminary Dissertation Concerning the Fundamental Principle of Virtue or Morality," p. xli.

55. Gay, "Preliminary Dissertation Concerning the Fundamental Principle of Virtue or Morality," p. xli.

56. Some of Gay's contemporaries doubted that our obligations must align with our happiness. For discussion, see Darwall, "Norm and Normativity."

57. Some of Gay's contemporaries also had doubts about the correspondence between God's sanctions and the merit or demerit of our actions. See, for example, Carmichael, *Natural Rights on the Threshold of the Scottish Enlightenment*, p. 347 and pp. 349–350.

58. Schneewind, *The Invention of Autonomy*, p. 404.

59. Sidgwick, *Miscellaneous Essays and Addresses*, p. 151.

60. For discussion see Wootton, "Helvétius."

61. For Helvétius' Epicureanism, see Force, "Helvétius as an Epicurean Political Theorist."

62. It is interesting to consider what supports claims of human equality when God is not invoked. For relevant discussion, see Waldron, *God, Locke, and Equality*.

63. Helvétius, *De l'esprit*, p. 41.

64. Helvétius, *De l'esprit*, p. 39.

65. Helvétius, *De l'esprit*, p. 70.

66. Helvétius, *De l'esprit*, p. 161. Compare to the opening of Bentham's *IPML*.

67. Helvétius, *De l'esprit*, p. 81.

68. Helvétius, *De l'esprit*, p. 37n.

69. Helvétius, *De l'esprit*, p. 189 and p. 190.

2 Bentham and utilitarianism in the early nineteenth century

James E. Crimmins

Utilitarian praxis

Jeremy Bentham (1748–1832) coined the term 'utilitarian' in the summer of 1781, when he recorded a dream in which he "was the founder of a sect; of course a personage of great sanctity and importance. It was called the sect of the *utilitarians*."[1] The dream turns on Bentham's hopes for *An Introduction to the Principles of Morals and Legislation* (*IPML*), printed the previous year (but not published until 1789), "my driest of all dry metaphysics,"[2] parts of which he had recently read to the company of guests at the country seat of his patron, the reformist Whig the Earl of Shelburne, who served as Prime Minister 1782–1783 and became Marquis of Lansdowne in 1784. In Bentham's telling of the dream he writes, "there came to me a great man named L. [Shelburne] and he said unto me, what shall I do to . . . save the nation? I said unto him – take up my book, & follow me." With the noble lord in tow, he then encountered King George III and instructed his "apostle," Shelburne, to give the king "a page of my book that he may read mark learn and inwardly digest it." Bentham's fanciful reverie is indicative of his strategy at this time for the implementation of utilitarian ideas. Inexplicably, however, he demurred from presenting Catherine the Great with a copy of *IPML* when the opportunity arose during his stay in Russia with his brother Samuel in 1787.[3] Nevertheless, the global reach of Bentham's ambitions is clearly signaled.[4] Its foundations lay in the "universal" character of the constructs of his secular version of utilitarianism. Allowing for the influence of local contexts in shaping laws based on utilitarian principles,[5] in *Of the Limits of the Penal Branch of Jurisprudence*, written 1780–1782, he outlined a system of "universal jurisprudence" premised on "the universal system of human actions"; in the preface to *IPML*, he detailed the several branches of a complete code of law or *pannomion*.[6] At his most universal Bentham, who was responsible for introducing the word "international" into both the French and English languages, held (in his "Principles of International Law," written from the perspective of "a citizen of the world" in 1786–1789) that the object of international law is "the greatest happiness of all nations taken together."[7] Consistent with this objective, in the later *Codification Proposal, Addressed . . . to All Nations Professing Liberal Opinions* (1822)[8] he advertised his credentials as a codifier of law to statesmen around the world, first by setting down the utilitarian principles of an "all-comprehensive code of law, accompanied with a perpetually interwoven rationale, drawn from the *greatest happiness*

principle,"[9] and second by providing testimonials to his aptitude for the task of codification. The testimonials came from far and wide between the years 1814 and 1822 – from fellow reformers in England, government ministers, and representatives of the Cortes in Spain and Portugal, Italian and French liberals, state governors, legislators, and other public figures in the "Anglo-American States," and Catherine's grandson Tsar Alexander, among others. Little came of the initiative, but it highlights a key feature of Bentham's utilitarianism: the unity of theory and practice – the "praxis" that embraced his philosophical engagement with the world.[10]

The "fundamental axiom" of the theory was first stated in *A Fragment on Government* (1776), where he states, "it is the greatest happiness of the greatest number that is the measure of right and wrong," and "the obligation to minister to general happiness, was an obligation paramount to and inclusive of every other."[11] As he later explained, the utility principle or greatest happiness principle "gives character and direction to the details of Morals and Politics . . . Government and Legislation . . . each of them considered *as it is*, for the hope of seeing it rendered what it *ought to be*."[12] It was in *IPML*, however, that he delineated the component parts of the "science" of morals and legislation founded upon the utility principle, and expressly laid it down as the objective of government "to rear the fabric of felicity by the hands of reason and of law."[13]

Bentham's theory of motives provided the bedrock of his utilitarianism. Pleasures and pains, he announced in the famous opening passage of *IPML*, are mankind's "sovereign masters." As the "real entities" of individual experience, acting both as the *final* cause of individual action, and as the *efficient* cause and means to individual happiness, they determine what people do and what they ought to do.[14] Pleasure, excepting the "immunity from pain," is "the only good," while pain "without exception, the only evil."[15] All motives have their source in the anticipation of pleasure or pain, and as such are properly termed pleasures and pains "in prospect." In this sense, interest is composed entirely of imagined (though not imaginary) expectations and apprehensions about the future. It is not the pleasures presently enjoyed or the pains presently suffered that provide the motive to action (though past experience will often influence a person's expectations of the future), but rather the belief or persuasion that the imagined outcome of an action will come to pass.[16]

In Bentham's account in *IPML*, all motives, including the most extensive benevolence, are rooted in self-interest; in these terms, the notion of a *dis*interested motive is implausible. People commonly speak of actions as proceeding from good or bad motives, but the expression is inaccurate. Even "ill-will" is still a kind of pleasure or pleasure in prospect that constitutes a person's motive. Does this mean there is no genuine philanthropy or self-sacrifice in the world? Bentham recognized the possibility of altruistic actions and frequently alluded to his own philanthropy when recommending

schemes that would further the public good. Moreover, he acknowledged that sympathy for others was a "primeval and constant source" of pleasure and action.[17] However, he still maintained that no action strictly speaking could properly be considered disinterested, since all action is "caused" by the anticipated pleasures and pains that constitute an individual's perception of his interest. Given the neutrality of motives, the utility of an act – its goodness or badness, rightness or wrongness – is based entirely on its consequences: the benefits and/or costs that result.

When deciding whether to act or which act to undertake a person must calculate as best he can the pleasures and pains that may reasonably be expected to accrue to the persons (including himself) affected by the acts under consideration. According to what others have called the "felicific calculus," the value of a pleasure or pain will be determined by its "intensity," "duration," "certainty or uncertainty," and its "propinquity or remoteness." Where the object is to measure the value of a pleasure or pain in terms of the tendency of an act, there are two additional circumstances to be taken into account: "fecundity," that is "the chance it has of being followed by sensations of the *same* kind," and "purity," or "the chance it has of *not* being followed by sensations of the *opposite* kind." Where there are a number of persons with reference to whom the value of a pleasure or a pain is considered, then the "extent," or the number of persons affected, must be factored into the calculus.[18] Though Bentham believed there was nothing in what he had proposed "but what the practice of mankind, wheresoever they have a clear view of their own interest, is perfectly conformable to," it has often been said that applying the felicific calculus is impractical. But Bentham realized that neither the individual nor the legislator could strictly follow the process he described. Rather, he presented it as a model of an ideal calculation, and insisted that "as near as the process actually pursued on these occasions approaches to it, so near will such process approach to the character of an exact one."[19]

Clearly, Bentham was aware of the limitations of the mathematical approach to summing pleasures and pains. As recent scholars have noted,[20] his classification of pleasures included qualitative distinctions not amenable to strict calculation. It is impossible, for example, to quantify the intensity or purity of a pleasure. On the other hand, it is entirely feasible for an individual to determine that one pleasure is more intense or purer than another he has experienced and to quantify multiple qualities of pleasures, though Bentham understood that such "calculations" were more impressionistic than mathematical.[21] The complexity inherent in the idea that pleasures could be estimated by anyone who had "a clear view of their own interest" underscores Bentham's instruction that great care is needed in designing institutions and crafting laws. The development of an objective theory of morals and politics was intended to guide the moralist, educator, and legislator in the assessment of interests and the sorts of cost–benefit analysis that could, potentially, produce the optimum utility for all those affected by the decisions of those in authority.[22] Viewed in this light, the distance between

Bentham and the supposed "revisionism" of Mill's distinction between higher and lower pleasures is sharply reduced.

Based on the calculation of interests, the goal of the legislator is to enhance the greatest happiness of the community by formulating laws aimed at maximizing the happiness of the particular individuals who make up the community. This is "the sole end which the legislator ought to have in view" and "the sole standard in conformity to which each individual ought, as far as depends upon the legislator, to be *made* to fashion his behaviour."[23] Properly constructed laws, with the attendant rewards and sanctions (physical, political, moral, and religious) that give them their binding effect,[24] both reflect the interests of the people and construct interests by providing individuals with the motives to pursue courses of action beneficial to the community. This is most obviously the case in penal law, where punishment is designed to deter individuals from actions harmful to the interests of others, but is also applicable across all the other branches of the law. In demonstrating how best this could be done, the task Bentham set for himself was to map out the rules, derivative precepts, maxims, and subordinate principles to guide the legislator in crafting penal, civil, constitutional, and procedural law, and their various subsets.[25] It was a huge undertaking that dominated extensive periods of his life right up until his final years, when, at the age of 82, he assured a correspondent he was still "codifying like a dragon."[26]

However, Bentham realized that spelling out the foundations of his philosophy and laying down explicit guidelines for the implementation of the utility principle in each branch of the law would not ensure their adoption. Nor were appeals to those in power a reliable route to success. The frustrating engagement with the Pitt administration over the panopticon in the 1790s put a severe dent in his expectations from this quarter, even if he continued to think that those who had acquired power in revolutionary France, republican America, and other states wrestling with new constitutions would be more amenable to his overtures.

Bentham's *Panopticon; Or the Inspection House* (1791) has occasioned a good deal of controversy in discussions of the history of punishment.[27] The basic architectural idea came from Bentham's brother Samuel, a naval architect in the service of the Russian Prince Potemkin. Jeremy immediately saw its potential for adoption in any institutional context requiring a high level of supervision, including schools, hospitals, factories, and poorhouses. It is, however, as the governing principle of a prison, in which convicted criminals would be subject to a disciplinary regime based on the maxim that "the more strictly we are watched, the better we behave,"[28] that it has gained most attention. The circular prison design was to leave each cell visible to the center which was occupied by a watchtower from where the unseen warden might observe the activities of prisoners day and night. The real achievement of the idea, however, was the manner in which Bentham gave effect to utility in a range of practical subordinate principles: economy, since the prison should be a private self-sustaining operation not requiring financial

assistance from the public purse; severity, because it was necessary for the offender to suffer to serve the ends of reformation and deterrence; and humanity, which prescribed that prisoners should be deprived only of liberty, not of health or life. In contrast with the cesspits of the existing gaols and hulks in Britain, and the horrific experiment with the penal colony at Botany Bay, Bentham's prisoners were to be kept clean and their labor was to be productive and profitable, and serve to develop skills that might be useful to them upon release and assist in their moral reformation. In support of these objectives, Bentham devised several devices to produce transparency and accountability. The chief mechanism intended to bring the interest of the warden in line with his duty to be humane was publicity – "the most effectual means of applying the force of moral motives, in a direction tending to strengthen the union between his interest and the *humane* branch of his duty; by bringing to light, and thus exposing to the censure of the law and of public opinion . . . every instance of contravention."[29] Members of Parliament and interested members of the public were to be guaranteed free access to the prison, making the panopticon subject to "the great *open committee* of the tribunal of the world."[30] The aim was to prevent abuses of power by prison officials and to enhance the security of the inmate.

Bentham's hopes for the panopticon were never realized, despite the Pitt administration entering into a contract to enable him to establish and manage a penitentiary in London under the terms of the Penitentiary Act 1794. Collusion between aristocratic landowners and public officials stymied the project; Bentham even suspected George III of meddling behind the scenes.[31] After many years of negotiations, lobbying, and repeated disappointments, he was forced to admit defeat in 1802, and ten years later Parliament voted £20,000 in compensation for Bentham's immense investment of time, money, and effort. The entire experience left him bitter about the motives of those in public life, and served to confirm in him the notion that democratic institutional arrangements were necessary to provide "securities against misrule."

Long before then, however, Bentham had realized the need to broaden the base of supporters for his ideas – men willing to disseminate and prosecute his ideas and press for reform based on utilitarian principles. His own writings – with the notable exceptions of *A Fragment on Government* and *Defence of Usury* (1787) – were notorious for their tortuous terminology and distracting parentheses, castigated by the critics as incomprehensible "Benthamese," making them ill-suited for public consumption. In rendering utilitarian ideas in a readable form, no one was more important than the Genevan lawyer Pierre-Étienne-Louis Dumont, who assumed the role of Bentham's translator and editor, and fashioned an international audience for his work.[32]

The dissemination and reception of utilitarianism

Dumont become acquainted with Bentham in April 1788 in the circle of reformers

connected to Lord Lansdowne, and soon after gave his assistance in correcting several essays the philosopher had penned in French in response to events in France.[33] He then translated *Panopticon* for distribution to the members of the French National Assembly,[34] and in the summer of 1792 embarked upon the formidable challenge of composing a work that would present Bentham's legal philosophy in a comprehensive and accessible form. Between October 1796 and April 1798 Dumont published a series of extracts from this ongoing work (and from Bentham's *Manual of Political Economy*) in the Genevan journal the *Bibliothèque britannique*.[35] The complete work appeared in three volumes in 1802 as *Traités de législation civile et pénale* (1802), the main parts of which were the treatises on civil and penal law. Four more major publications based on Bentham's writings followed: *Théorie des peines et des récompenses* (1811), *Tactiques des assemblée législatives, suivi d'un traité des sophismes politiques* (1816), *Traité des preuves judiciare* (1823), and *De l'organisation judiciaire et de la codification* (1828). None of these texts, including the *Traités*, was a straightforward translation of Bentham. They are commonly called redactions or recensions, either edited from Bentham's original papers or combining manuscript material with extracts from published works, substantially rewritten in plainer language and divested of cumbersome detail. In this form utilitarian ideas reached many more readers than Bentham ever could have hoped. True, *A Fragment on Government* had caused a minor stir, but *IPML* had virtually sunk without trace when it appeared in 1789 – "the edition was very small and half of that devoured by the rats," he noted.[36] Matters were far otherwise with the *Traités*, substantial parts of which were based on *IPML*.[37]

Though questions have been asked about the faithfulness of Dumont's publications to the original,[38] the *Traités* is of unquestionable importance in fathoming the nature of Bentham's philosophy as it would have appeared to the world in the first half of the nineteenth century. It proved to be one of the great works of legal philosophy in an age that could boast many other splendid examinations of the principles and forms of law. Even the young John Stuart Mill – so close to the fountainhead of the doctrine we might have expected his confirmation to have been initiated by an immersion in the original script – cut his utilitarian teeth on Dumont's volumes, professing that the *Traités* marked "an epoch in my life." "I now had opinions," he declared, "a creed, a doctrine, a philosophy; in one among the best senses of the word, a religion; the inculcation and diffusion of which could be made the principal outward purpose of a life."[39]

Dumont records that 3,000 copies of the *Traités* were initially distributed in France, and that it was "frequently quoted in many official compositions relating to civil or criminal codes."[40] Soon after, it was translated into Russian, and later into Spanish, German, Hungarian, Polish, and Portuguese.[41] Other editions of the *Traités* followed. Reportedly, 50,000 copies of Dumont's various recensions were sold in Europe in the early decades of the century and 40,000 in Spanish translation in Latin America alone.[42]

Bentham's ideas had been circulating in Spain since 1810 through the London-based *El Español*,[43] but interest increased during the liberal triennium of 1820–1823. In 1820, Toribio Núñez, librarian at the University of Salamanca, published a two-volume account of utilitarian legal philosophy based on the *Traités*, and his own work on moral and political philosophy was greatly influenced by Bentham's ideas.[44] In 1821–1822, Ramón de Salas, a law professor at Salamanca, produced the first Spanish translation of the *Traités* in five volumes,[45] which received a scathing critique by José Vidal, a Dominican theologian at the University of Valencia, who condemned the work as an encouragement to revolution.[46] From Spain, Bentham's utilitarianism reached its former colonies in the New World. Andrés Bello used Salas' translation as the basic text for his law lectures at the Colegio de Santiago in Chile, as did Pedro Alcántra de Somellera, professor of civil law at the University of Buenos Aires.[47] In 1825 Francisco de Paula Santander, the Vice-President of Gran Colombia, decreed that the work be required reading for all law students in the vast territories of the new republic, but in 1828 its President Simón Bolívar, the legendary "Liberator," after previously embracing the principles and purpose of Bentham's legal philosophy, bowed to clerical pressure and banned its teaching as detrimental to religion, morality, and social order.[48] Santander, more inclined to resist the influence of the Catholic Church, restored it to the curriculum of the universities when he became President of the newly constituted state of Colombia in 1832.

Bentham's ideas were also beginning to take hold in other parts of the world. Following the Greek revolution against Ottoman rule, the historian and legal scholar Anastasios Polyzoides, who had a hand in drafting its new constitution in 1822, translated an extract on "publicity" from Dumont's *Tactiques des assemblée legislatives* for a Missolonghi newspaper in 1824, promoting transparency in legislative proceedings and government matters in general. A year later he published *A General Theory of Administrative Systems and especially of the Parliamentary One, Accompanied by a Short Treatise on Justices of the Peace and Juries in England* (1825), containing a defense of representative government and advocating a judicial system based on utilitarian principles, replete with references to Bentham.[49]

In the United States, the dissemination of utilitarianism was initially hampered by the absence of an English translation of the *Traités*, but there too Bentham's influence was not long in being felt. Utilitarian ideas were first introduced into the academic study of the law by David Hoffman, the inaugural professor of law at the University of Maryland law school, which he helped found in 1816. John Neal, who studied law under Hoffman's guidance, described him as one of Bentham's "most enthusiastic admirers."[50] Hoffman's bibliographic *A Course of Legal Study* (1817), later expanded to a two-volume edition (1836), became a standard guide for the teaching of law in American universities, and continued to hold its place of eminence in the field well into the second half of the century.[51] In the published version of his lectures, *Legal Outlines* (1829), he expounded a legal theory that combined utilitarianism with elements of natural law. In *A*

Course of Legal Study Hoffman recommended that students study closely the first seven chapters of *IPML* and chapters 12–17 of Dumont's redaction *Théorie des peines et des récompenses*. Of the latter he wrote: "It is a matter of no less surprise than regret that a work of such extraordinary merit . . . should so long have continued unknown, not only to the students, but to the learned of our country." Nor did he think this was putting it too strongly, for "nowhere among ancient or modern productions, is the philosophy of criminal legislation so ably and happily illustrated." In this regard Bentham "left his predecessors at an immeasurable distance."[52]

Hoffman encouraged Neal to translate the *Traités* into English,[53] a task he undertook during the eighteen months he stayed with Bentham in London in 1825–1826. However, Neal managed to translate only the introductory "Principes de legislation" and balked at the essays on civil and penal law. His edition of the *Principles of Legislation* first appeared in the United States in parts in *The Yankee*, which he edited under the banner heading "the greatest happiness of the greatest number," then in full in 1830. In the biographical reflections on Bentham included in the volume Neal described the philosopher as "the great high-priest of legislation" and commented in effusive terms on Dumont's role in popularizing his work.[54] Neal reported that four hundred of the five hundred printed copies were sold soon after publication.[55]

It was Richard Smith, a government tax officer and one of Bentham's young disciples, who eventually translated the civil and penal law parts of the *Traités* into English for the *Works of Jeremy Bentham* in 1838.[56] While Smith was at work in England, the historian and anti-slave propagandist Richard Hildreth was busy translating the same material on the other side of the Atlantic, convinced that the widespread interest in legal reform in the United States would benefit enormously from Bentham's ideas. Hildreth's translation eventually appeared in two volumes in 1840 as *Theory of Legislation*, and remained at the center of utilitarian studies in the English-speaking world through to the middle of the twentieth century.[57] Two reviews appeared in the American journals of the day, both of which praised Bentham's legal philosophy, while objecting to its underlying moral assumptions and disdain for religion,[58] a position frequently adopted in contemporary appraisals of utilitarianism. The reviewers paid particular attention to the systematic presentation of Bentham's theory of civil law – notably his delineation of the subordinate ends of civil law (security, subsistence, abundance, and equality) in relation to property – which had appeared for the first time in print in the *Traités*, and which provided the guiding principles for his writings on the poor laws and economic policy.[59] Dumont himself believed that Bentham's theory of civil law was one of the most significant of his contributions to legal philosophy,[60] and it was this feature of his work that most impressed itself on the teaching of law in the newly independent states of South America, where property rights were a matter of considerable import in the aftermath of the collapse of the Spanish and Portuguese empires. Sir James Fitzjames Stephen, the English jurist and critic of Mill's *On Liberty*, later commented in a review of Hildreth's

1864 edition that Bentham had single-handedly rescued the theory of civil law from undeserved neglect.[61]

Hildreth, whose own *Theory of Morals* (1844) drew substantially on the *Traités*, was correct in thinking that law reformers in the United States would find sustenance in Bentham's legal philosophy. Thomas Cooper, who left England for the United States in 1794 with Joseph Priestley, from whom he initially derived his utilitarianism, by the 1820s was a confirmed Benthamite and the recipient of writings personally sent by the philosopher.[62] Thereafter, with the exception of the slavery question,[63] he systematically employed utilitarian principles in his writings on political economy and law. Edward Livingston, the famous author of codes of law for Louisiana,[64] freely professed that his codification work was shaped by Bentham, whom he acknowledged as the world's leader in this domain of law,[65] and confirmed that it was his reading of the *Traités* which "fortified me in a design to prosecute the subject" of the reform of penal law.[66] Gilbert Vale used his position as editor of the radical periodical *The Diamond* (1840–1842) – like *The Yankee*, published with the banner heading "the greatest happiness of the greatest number" – to disseminate Bentham's criticisms of the vagaries, chicanery, and technicalities of the law. Inspired by Bentham, Vale was in favor of humane penal laws that proportioned penalties to the objective of deterrence and an advocate of state intervention to alter the social circumstances which fostered crime.[67] John O'Sullivan, the quixotic editor of the *United States Magazine and Democratic Review* famous for coining the phrase "manifest destiny," wrote sympathetic reviews of Bentham, Livingston, and Hildreth, and followed them in advocating utilitarian law reform, particularly the abolition of capital punishment.[68]

Bentham's influence continued throughout the century in America, where the *Traités* paved the way for the reception of other editions and versions of his writings, which in turn led to sympathetic responses to the more amenable forms of utilitarian moral and legal theory offered by Austin, Mill, and Sidgwick.[69] On the other hand, Bentham's *political* prescriptions made little impact in the United States, which was, by comparison to aristocratic England, already an advanced democracy. If the utilitarian *Constitutional Code* was directed "for the use of all nations and all governments professing liberal opinions," as its title page declared, the political positions it embraced were recommended, in the first instance, for adoption at home.

Utilitarian politics in Britain

For all Bentham's success abroad, in the early years of the new century he was little known in his own country, save among a small band of law reformers determined to tackle the antiquated and notoriously harsh punishments meted out by English penal law. His reputation, such as it was, fueled the caricatures of Lord Byron, who voiced doubts

about the balance of his mind, and William Hazlitt, who depicted him as a venerable anchorite in the quiet of his cell reducing law to a system and the mind of man to a machine, divorced from the life of spirit, imagination, passion, and sentiments of love, and who dismissed utilitarianism as a philosophy "fit neither for man nor beast."[70] Bentham, Hazlitt fancifully opined, was the head of a zealous sect of "philosophical projectors . . . churning out inventions in jurisprudence, morals, logic, political economy, and constitutions, with as little variation as a barrel-organ plays 'God save the King' or 'Rule Britannia'."[71] There were times when Bentham was bitter about his lot in life – let down by Shelburne from whom he impetuously demanded a seat in the House of Commons, humiliated by a government that failed to keep its word to support the building of a panopticon penitentiary, and his philosophical contributions to understanding and reforming law for many years met by either bemusement or indifference. He struck a note of despondency when he remarked how well his work had been received in France: "Greater – far greater – is the honour bestowed upon him in that foreign country than in his own," he wrote.[72] Even those unconvinced of the merits of utilitarianism recognized the irony of the situation. Hazlitt, who was for a time a tenant of Bentham's, with only a little extravagance opened an 1824 essay on his former landlord with the declaration:

> Mr. Bentham is one of those persons who verify the old adage, that "A prophet has most honour out of his own country." His reputation lies at the circumference; and the lights of his understanding are reflected, with increasing lustre, on the other side of the globe. His name is little known in England, better in Europe, best of all in the plains of Chili [*sic*] and the mines of Mexico.[73]

The critic William Empson concurred. "Mr Bentham's reputation," he wrote, "is at present thoroughly European . . . he has been left almost 'a stranger in his father's house'."[74]

Hazlitt and Empson were only half right. In the years following the defeat of Napoleon in 1815, when calls for legal, social, and political reform gained considerable traction, Bentham's reputation in Britain underwent a change from that of a misunderstood oracle on the periphery of the intellectual and political world to that of a venerable sage situated at the center of a broad reform movement. An increasingly impressive list of public figures cited his authority to enhance their credibility both within and outside Parliament, including law reformers such as Samuel Romilly, James Mackintosh, and Henry Brougham, and political radicals like Francis Burdett and Daniel O'Connell. Bentham sought to extend this influence to include populist agitators, like John Cartwright, William Cobbett, and Henry Hunt, but with less success. These were men with their own supporters and agendas, and they did not often see eye to eye, and could not be expected to work harmoniously with the kind of reformers with whom Bentham liked to dine and discuss the issues of the day – the likes of Romilly, Brougham, and Joseph Hume. Nor

were Bentham's utilitarian disciples entirely comfortable with the alliances he felt it necessary to construct to bring about reform.[75] Understandably, there were moments of tension and dispute between Bentham and his Whig friends. His intimacy with Romilly was sorely tested by the latter's stubborn moderation on political reform, which led Bentham to an egregious betrayal of his friend when he entered into a pact with Burdett aimed at preventing Romilly's election for Westminster in 1818. Brougham also frustrated Bentham by his half-hearted support for reform. He once announced in the Commons that, "The age of law reform and the age of Jeremy Bentham are one and the same,"[76] but this did not prevent Bentham from attacking the Lord Chancellor in *Lord Brougham Displayed* (1832), a criticism of Brougham's plan to absorb the courts of the Vice-Chancellor and the Master of the Rolls into the Chancery Court, a plan which fell far short of the reform of the judiciary Bentham advocated.[77]

The transformation in Bentham's reputation in Britain can be traced to the early months of 1809 when, with the encouragement of his newly enlisted aide-de-camp James Mill, he began writing extensive drafts of analysis on the forms and debilitating consequences of "influence" in British politics, material which formed the grounds for his support for representative democracy. In these still-unpublished papers Bentham explored the relation between abuses in the law and abuses in Parliament: the beneficiaries of the law and the beneficiaries of the corrupt Parliament were united in one "confederated sinister interest," he concluded.[78] The terminology of "sinister interest" had entered Bentham's vocabulary several years before as shorthand for his view that the interests of England's rulers in Church, law, and government were invariably pursued at the cost of the interests of the people, and hard evidence of corruption in government came to hand in 1809 in the reports of the Commons' Select Committee on Public Expenditure detailing the exorbitant amounts of public money disbursed in support of sinecure posts and pensions.[79] With the long-delayed publication of *Plan of Parliamentary Reform* in 1817,[80] Bentham formally announced the tenets of his democratic politics: the elimination of royal patronage, a substantial extension of the franchise, annual elections by secret ballot, the election of intellectually qualified independent Members of Parliament (with a system of fines to ensure regular attendance), and the accurate and regular publication of parliamentary debates. Collectively styled "securities against misrule," none was more important than the "Public Opinion Tribunal," the open court of public opinion founded on the freedom of the press, by which government actions could be held up to public scrutiny and officials held accountable. Much of what Bentham recommended was governed by "the *interest-junction-prescribing* principle" designed to ensure that the interests of those with power would be reconciled with the public interest.[81] The same reasoning had informed Bentham's strategy for holding the warden to account in the panopticon, but the basic political idea can be traced to the earlier *A Fragment on Government*, in which he argued that effective government required institutions and practices that enabled "the frequent

and easy *changes* of condition between govern*ors* and govern*ed*; whereby the interests of one class are more or less indistinguishably blended with those of the other."[82] By 1809, he became convinced that democratic institutions and procedures would be necessary to achieve this identification of interests. And when he turned his thoughts to constitutional law in earnest in the 1820s, by then fully committed to republicanism, it was with the conviction that all states in which the institutions of representative democracy already existed or in which they could be introduced were fertile soil for the utilitarian *pannomion*.

The response to Bentham's democratic proposals from apologists for Britain's "mixed and balanced" constitution was predictable. The Tory *Quarterly Review* disparaged the deconstruction of the foundations of the constitution, but without stooping to examine the merits of the reforms proposed.[83] The stiffest opposition came from Mackintosh in the *Edinburgh Review*,[84] the vehicle for moderate Whig reform. Mackintosh denounced Bentham's position on democratic reform as dangerous radicalism, arguing it went too far in extending the franchise – a plan that would endanger the liberties of the people, he thought – and placed unfounded faith in the beneficial effects of the secret ballot and annual parliaments. Like many reformist Whigs, he believed the extension of the franchise should be gradual and limited, with attention focused, initially at least, on the disfranchisement of rotten boroughs, and argued that the variety of franchises that ensured representation for different sections of the populace should be maintained.[85]

Bentham affected indifference to Mackintosh's attack, but Mill saw it as treachery by a fellow reformer. It seems likely that Mackintosh's critique was the catalyst for Mill's own statement of the political implications of utilitarian doctrine in the famous essay "On Government," published in the *Encyclopaedia Britannica* in 1820. Mill had made flattering remarks on Mackintosh's tenure as recorder for Bombay in the pages of his *History of British India* (1818), but after the attack on Bentham his correspondence contains only disdain whenever his fellow Scot came to mind.[86] Later, Thomas Babington Macaulay, exercised by the republication of Mill's essay in 1825, engaged in a six-part debate with the utilitarian camp – one of the period's great intellectual battles played out in the pages of the *Edinburgh Review*, in which Mackintosh had previously denounced Bentham's views, and the *Westminster Review*, the mouthpiece of Benthamite utilitarianism.[87]

In the essay "On Government" Mill attempted to distill the essence of the utilitarian position on political reform. Though lacking the penetration of Bentham's critique of established institutions and exhibiting little of the senior utilitarian's subtlety in linking theory to practice, it had the virtue of being a systematic and forthright exposition based on the "science of human nature." In determining the extent of the suffrage, however, an element of caution entered Mill's argument which was unwarranted by his initial premises and which exposed him to the criticisms of the more progressive among his fellow radicals. The franchise, he argued, might be limited by excluding "all those individuals

whose interests are indisputably included in those of other individuals," i.e., children "up to a certain age," since their "interests are involved in those of their parents"; women, "the interest of almost all of whom is involved either in that of their fathers or in that of their husbands"; and men under the age of 40, on the grounds that older males could be counted on to possess "a deep interest in the welfare of the younger men." Mill thought a property qualification might be added to the age qualification, set at a level that encompasses the majority of men 40 and over and sufficient, therefore, to provide "a tolerable security" to good government.[88]

Mill's exclusions from the vote were not well received among the Benthamite radicals. Bentham was committed to universal manhood suffrage (subject only to a literacy qualification), and in notes he shared with Mill he disparaged the arguments by which his fellow utilitarian had justified limiting the vote to men of property over 40.[89] If Bentham was prepared to accept the politics of excluding women, at least temporarily, John Stuart Mill felt no such compulsion. He later remarked that the argument for denying women the vote was the worst passage his father ever wrote.[90] William Thompson, a socialist utilitarian and friend to Bentham and the younger Mill, shared their dismay. In *Appeal of One Half of the Human Race Women against the Pretensions of the other Half Men* (1825), he supported the general utilitarian position on political reform but offered a considered, if at times intemperate, critique of Mill's exclusion of women from the franchise. So incensed was he that he went so far as to demonstrate the inconsistencies inherent in the basic assumptions of Mill's theory, concluding that the exclusion of women was a "disgrace [to] the principle of utility."[91]

Macaulay, like Mackintosh, was sympathetic to the critical intent of utilitarian legal theory, but like Thompson he objected to the narrow view of self-interested human nature that underpinned Mill's rendering of the doctrine. Later, he regretted the tone of his critique of Mill, acknowledging that he might have "abstained from using contemptuous language respecting . . . [the author of] on the whole, the greatest historical work which has appeared in our language since that of Gibbon."[92] Nor was he averse to consulting Mill when drafting a penal code for India, and his position on education has been described as "James Mill's philosophy expressed in Macaulayese."[93] Politically, however, the central issue for Macaulay, as it was for many Whigs, was how to maintain support for moderate parliamentary reform while dismissing the position on manhood suffrage taken by Bentham and his supporters. Macaulay's approach was to demonstrate the defects in the foundations of utilitarian philosophy itself – at least, as they were presented in Mill's essay.

Macaulay's main criticism of Mill is that his way of proceeding is entirely *a priori*, by which he means that "certain propensities of human nature are assumed" and "from these premises the whole science of politics is synthetically deduced," a deduction which it is logically "utterly impossible" to draw. Added to this, the claim that men always act out of self-interest is at best a comment on only "one-half of human nature"[94] (a

criticism the younger Mill would later direct at Bentham)[95] and supposed that "the motives which impel men to oppress and despoil others" were the only motives influencing their actions.[96] On the basis of this false assertion Mill derived doctrines which are also false,[97] including the need for democratic checks on self-interested politicians and the implicit faith that only with the establishment of representative democracy could the public good be served.[98] If legislation is the primary means of constraining individuals, as Mill implied, and legislators act only out of self-interest, as do all men, then the attempt to produce an identification of interests between rulers and people through democratic institutions is futile: the principle of utility would only be operable by a happy coincidence.

Much of Macaulay's line of attack on Mill was reiterated by Mackintosh a year later in "Dissertation on the Progress of Ethical Philosophy" (1830), an extended essay prefixed to the *Encyclopaedia Britannica*, in which, contrary to William Paley and Bentham, he maintained that ethics rested on the primacy of conscience. Mackintosh was particularly harsh in his observations on Mill.[99] Mill's unmitigated response is contained in *A Fragment on Mackintosh* (1835),[100] in which he restated the principal parts of his original position in "On Government," insisted that institutions designed to produce an identity of interests between the governors and the governed are "the only security for good government," and left his original exceptions to the ballot intact.[101]

Conclusion

Bentham's role in the great debate was not significant.[102] Initially, Perronet Thompson, the newly minted proprietor of the *Westminster Review*, had hoped that the response to Macaulay would come from Bentham. Once again, however, he left the debate to others, supplying only notes on the history of the utility principle for Thompson to use in crafting a defense.[103] Enmeshed in the politics of constructing a broad alliance of reformers under his leadership, Bentham had bigger fish to fry. By the late 1820s his hopes were high for law reform and political reform and he was anxious to avoid friction within the loose alliance of moderate and radical reformers he had worked so hard to construct. Moreover, momentous political events abroad had encouraged him to press on with codifying the political, administrative, and judicial institutions, rules, and practices that should be in force in all nations professing liberal opinions. Ultimately, Bentham's attempt to fulfill the role assigned him by the Guatemalan politician José del Valle – "Legislador del mundo"[104] – fell short of its initial promise. Nevertheless, this was not the man described in the *Quarterly Review* whose litany of failed attempts "to become the governor of a prison, the enlightener of the world, the legislator of despotic Russia, of republican America, and lastly the head of a chrestomathic school," cast "a misanthropical gloom over his temper."[105] Rather this was the energized reformer who

believed that the most important task he could undertake was to produce a comprehensive code of law or set of codes based on the greatest happiness principle that would serve as a blueprint for reformers at home, and which might be exported piecemeal or in whole to other parts of the world.[106] To this end Bentham fostered friendships with prominent politicians, intellectuals, and reformers around the world. In the United States this included presidents Madison, Quincy Adams, and Jackson, and a variety of state governors and other legislators to whom he issued a standing offer to assist in law reform and sent batches of his writings to stimulate their interest. He used his connections to offer his services in drafting or codifying law to the Spanish Cortes in the aftermath of the restoration of the liberal constitution in 1820, to the Portuguese Cortes when it instituted a new constitution based on the Spanish example in 1821, to Tripoli when it appeared on the verge of revolution in 1822, to Greece following the promulgation of its new constitution in 1822 and the legislative debates on further reform 1823–1824, and to several of the newly sovereign states of Central and South America emerging from under colonial rule.[107]

John Bowring, a constant companion during these years and later the editor of the first collected edition of Bentham's writings, remarked "These were days of boundless happiness to Bentham, when, from every side, testimonials of respect and affection were flowing towards him, and when all events seemed concurring in advancing the great interest to which he was devoted."[108] Westminster was the hub of radical politics in England, and Bentham's home in Queen Square Place became a place of pilgrimage for disciples from near and far. Inspired, encouraged, flattered, and paid the most gratifying respects, as he was, none of his codification proposals came to fruition. Nevertheless, the news that his ideas were reaching all parts of the globe gave impetus to his resolve to complete the drafting of the constitutional code, the centerpiece of the utilitarian *pannomion*. The author of this project was by then an elder statesman in the world of legal philosophy and political radicalism, a man respected as the leader of the utilitarian school whose philosophy would occupy a central place in the discussion of legal and political thought and continue to inspire reform in Britain and elsewhere for much of the remainder of the century.

I am grateful to the editors for their meticulous reading and many helpful suggestions for improvements to this chapter.

Notes

1. Bentham, *Papers*, box CLXIX, fol. 79. For the complete text, see Crimmins, *Secular Utilitarianism*, pp. 314–316.

2. Bentham, *Correspondence*, vol. III, p. 69.

3. Bentham, *Correspondence*, vol. III, pp. 525–526.

4. I am indebted to David Armitage for the insights in "Globalizing Jeremy Bentham."

5. Bentham, *Works*, vol. I, pp. 169–194.

6. Bentham, *Of the Limits of the Penal Branch of Jurisprudence*, p. 17 and p. 130; and Bentham, *IPML*, p. 6, p. 8, and p. 305.

7. Bentham, *Works*, vol. II, p. 538.

8. Bentham, *"Legislator of the World,"* pp. 241–384; see also Lieberman, "Bentham on Codification."

9. Bentham, *"Legislator of the World,"* p. 260.

10. Bentham, *Papers*, box XCVII, fol. 5: "Obstacles Prejuges Professional – against Theory X Practise"; Bentham, *Church-of-Englandism and its Catechism Examined*, p. 373n.

11. Bentham, *Comment and Fragment*, p. 393 and pp. 440–441.

12. Bentham, "Article on Utilitarianism," in *Deontology*, p. 318.

13. Bentham, *IPML*, p. 11.

14. Bentham, *IPML*, p. 11.

15. Bentham, *IPML*, p. 100.

16. See Engelmann, "Imagining Interest."

17. Bentham, *IPML*, p. 61.

18. Bentham, *IPML*, pp. 38–39.

19. Bentham, *IPML*, p. 40.

20. Warke, "Multi-Dimensional Utility and the Index Number Problem"; and Rosen, *Classical Utilitarianism from Hume to Mill*, pp. 174–180.

21. Rosen, *Classical Utilitarianism from Hume to Mill*, pp. 177–178.

22. Rosen, *Classical Utilitarianism from Hume to Mill*, p. 179.

23. Bentham, *IPML*, p. 34.

24. Bentham, *IPML*, p. 34.

25. For the details, see Crimmins, *Utilitarian Philosophy and Politics*, chapter 4.

26. Bentham, *Works*, vol. XI, p. 33.

27. See, for example, Foucault, *Discipline and Punish*, and Ignatieff, *A Just Measure of Pain*. For a more balanced account, see Semple, *Bentham's Prison*, and Blamires, *The French Revolution and the Creation of Benthamism*, chapters 1–2.

28. Bentham, *Papers*, box CLII, fols. 332–333.

29. Bentham, *Works*, vol. VIII, p. 380.

30. Bentham, *Works*, vol. IV, p. 46.

31. "History of the War between Jeremy Bentham and George III, By one of the

Belligerents," in Bentham, *Works*, vol. XI, pp. 96–105.

32. Blamires, *The French Revolution and the Creation of Benthamism*, chapters 7–9.

33. Bentham, *Correspondence*, vol. VII, pp. 17–19.

34. Bentham, *Panoptique* (1791), reprinted in *Traités*, vol. III, pp. 201–272.

35. Bentham, *Correspondence*, vol. V, p. 200n.

36. Bentham, *Correspondence*, vol. IV, p. 34.

37. See the Comparative Table of Editions, in Bentham, *Theory of Legislation* [*Traités*], vol. I, pp. xlv–xlvi.

38. See Bentham, *Theory of Legislation* [*Traités*], introduction, pp. xii–xxi.

39. J. S. Mill, *Autobiography*, *Collected Works*, vol. I, pp. 67–68.

40. Bentham, *Works*, vol. I, p. 388.

41. Halévy, *The Growth of Philosophic Radicalism*, p. 530.

42. Bentham, *Works*, vol. XI, p. 33 and p. 88; see Avila-Martel, "The Influence of Bentham on the Teaching of Penal Law in Chile," McKennon, "Benthamism in Santander's Colombia," and Luño, "Jeremy Bentham and Legal Education."

43. Dinwiddy, "Bentham and the Early Nineteenth Century," p. 20.

44. Toribio Núñez, *Espiritu de Bentham ó sistéma de la ciencia social, ideado por Jeremías Bentham* (Madrid, 1820); *Principios de la ciencia social ó de las ciencias morales y politicas* (Salamanca, 1821). See Bentham, *Correspondence*, vol. X, pp. 329–337 and pp. 463–471.

45. *Tratados de legislación civil y penal*, 5 vols., ed. Ramón de Salas (Madrid, 1821–

1822).

46. José Vidal, *Orígen de los errores revolucionarios de Europe, y su remedio* (Valencia, 1827).

47. Pedro Alcántra de Somellera, *Principios de derecho civil* (Buenos Aires, 1824); see Bentham, *Correspondence*, vol. XI, p. 145n.

48. Bentham, *Works*, vol. X, pp. 552–554.

49. Peonidis, "Bentham and the Greek Revolution."

50. Neal, *Wandering Recollections*, p. 300.

51. King, *Utilitarian Jurisprudence in America*, p. 139.

52. Hoffman, *A Course of Legal Study*, 1817 edn., pp. 226–228.

53. Legaré, "Jeremy Bentham and the Utilitarians", pp. 267–268.

54. Neal, *Principles of Legislation*, p. 14 and pp. 10–11.

55. John Neal to Bentham (11 March 1830), *Correspondence*, vol. XIII (forthcoming).

56. Bentham, *Works*, vol. I, pp. 297–580.

57. C. K. Ogden's 1931 edition of *The Theory of Legislation* [*Traités*] is a reprint of Hildreth's 1864 edition.

58. "Jeremy Bentham" and "*Theory of Legislation*, by Jeremy Bentham."

59. Kelly, "Utilitarianism and Distributive Justice," pp. 62–81; and Quinn, "A Failure to Reconcile the Irreconcilable?" pp. 320–343.

60. Bentham, *Theory of Legislation* [*Traités*], vol. I, p. 91.

61. J. F. Stephen, *Horæ Sabbaticæ*, vol. III, p. 215.

62. Bentham, *Correspondence*, vol. IX, pp. 14–15.

63. Cooper, "Slavery."

64. Livingston, *A System of Penal Law for the State of Louisiana*.

65. Bentham, *Works*, vol. XI, p. 23 and p. 51. See also Bentham's letter to President Jackson, accompanying a copy of his *Papers Relative to Codification and Public Instruction* (1817) (*Works*, vol. XI, p. 40).

66. Bentham, *Works*, vol. XI, p. 23.

67. King, *Utilitarian Jurisprudence in America*, pp. 307–312.

68. O'Sullivan, *Report in Favor of the Abolition of the Punishment of Death by Law*.

69. Crimmins and Spencer, *Utilitarians and their Critics in America*, vol. I, introduction.

70. Hazlitt, "The New School of Reform: A Dialogue between a Rationalist and a Sentimentalist," in Hazlitt, *The Complete Works*, vol. XII, p. 184.

71. Hazlitt, "Sects and Parties," in Hazlitt, *The Complete Works*, vol. XII, p. 266.

72. Bentham, "Article on Utilitarianism," in *Deontology*, p. 311 and p. 312. See also Bentham, *Correspondence*, vol. V, p. 253.

73. Hazlitt, "Jeremy Bentham," in Hazlitt, *The Complete Works*, vol. XI, p. 5.

74. Empson, "Bentham's Rationale of Evidence," p. 458.

75. See Crimmins, *Utilitarian Philosophy and Politics*, chapter 7.

76. Radzinowicz, *A History of English Criminal Law and its Administration from 1750*, vol. I, p. 355.

77. Bentham, *Works*, vol. V, pp. 549–612.

78. Bentham, *Papers*, box CXXVI, fol. 304.

79. See Schofield, *Utility and Democracy*, pp. 137–139; and Hume, *Bentham and Bureaucracy*, pp. 175–178.

80. Bentham, *Works*, vol. III, pp. 433–557.

81. Bentham, *Comment and Fragment*, p. 515.

82. Bentham, *Comment and Fragment*, p. 485.

83. "*Plan of Parliamentary Reform* . . . by Jeremy Bentham" (*Quarterly Review*).

84. "*Plan of Parliamentary Reform* . . . by Jeremy Bentham" (*Edinburgh Review*).

85. "*Plan of Parliamentary Reform* . . . by Jeremy Bentham" (*Edinburgh Review*), pp. 175–176.

86. Thomas, *The Philosophic Radicals*, p. 125.

87. See Crimmins, *Utilitarian Philosophy and Politics*, chapter 1.

88. Lively and Rees, *Utilitarian Logic and Politics*, pp. 79–82.

89. Bentham, *Papers*, box XXXIV, fols. 302–303: "J. B. versus Mill."

90. J. S. Mill, *Autobiography*, *Collected Works*, vol. I, p. 98.

91. Thompson, *Appeal of One Half of the Human Race*, p. ix; see also p. 5, p. 7, pp. 27–30, pp. 44–45, and pp. 54–56.

92. Collini, Winch, and Burrow, *That Noble Science of Politics*, p. 110.

93. Forbes, "James Mill and India," p. 23.

94. Lively and Rees, *Utilitarian Logic and Politics*, pp. 124–125.

95. J. S. Mill, "Bentham," *Collected Works*, vol. X, pp. 92–94.

96. Lively and Rees, *Utilitarian Logic and Politics*, p. 108.

97. Lively and Rees, *Utilitarian Logic and Politics*, pp. 125–126.

98. Lively and Rees, *Utilitarian Logic and Politics*, p. 119.

99. Mackintosh, *Dissertation on the Progress of Ethical Philosophy*, pp. 236–264.

100. See the extracts in J. Mill, *Political Writings*, pp. 304–314.

101. J. Mill, *Political Writings*, p. 309.

102. Crimmins, *Utilitarian Philosophy and Politics*, chapter 2.

103. See Bentham, "Article on Utilitarianism," in *Deontology*, pp. 283–328.

104. Bentham, *Correspondence*, vol. XII, p. 217.

105. "*Church-of-Englandism and Its Catechism Examined* . . . by Jeremy Bentham," p. 169.

106. Bentham, *"Legislator of the World,"* p. 260.

107. See Crimmins, *Utilitarian Philosophy and Politics*, pp. 10–17.

108. Bentham, *Works*, vol. X, p. 539.

3 Mill and utilitarianism in the mid-nineteenth century

Henry R. West

From the perspective of the twenty-first century, the publication of John Stuart Mill's *Utilitarianism* in 1861 was the most important event in ethics in the mid-nineteenth century and one of the most important developments in the history of utilitarianism. Mill (1806–1873) was the greatest British philosopher of the nineteenth century, and his essay defended to a wide public the secular utilitarianism that had been founded two generations earlier by Jeremy Bentham (1748–1832). Written by a recognized great philosopher, it persuaded academic philosophers from his time on to take utilitarianism as a competitor to the views of Aristotle and Kant as one of the greatest traditions in ethics. A short book that could be read with apparent understanding by an ordinary person, it popularized utilitarianism and became the most widely read statement of utilitarianism from that time to the present. Because of its importance as a text in college courses in ethics and its extensive discussion in philosophical journals, a detailed analysis of it will be given later in this chapter.

Mill belonged to the utilitarian tradition founded by Bentham. The principle of utility, according to which the production of happiness and the elimination of unhappiness should be the standard for the judgment of right action and for the criticism of social, political, and legal institutions, was proposed by many writers in the eighteenth century, but it was Bentham who attempted to build a complete system of moral and legal philosophy upon that basis, and it was Bentham whose doctrine became the basis of a reform movement in the nineteenth century.

Mill was a direct heir of Bentham's philosophy. His father, James Mill (1773–1836), became acquainted with Bentham and became an exponent of the utilitarian philosophy in articles for journals and the *Encyclopaedia Britannica*, applying Benthamite principles to such subjects as government, education, liberty of the press, and colonial policy.[1] He also did much to define the policy of a group of reformers known as the "philosophical radicals," but he is most famous for the education that he gave to his son John Stuart.

John Stuart never attended school. Instead, he was rigorously tutored by his father, starting to learn Greek at age 3, Latin at age 8, and higher mathematics, economic theory, and nearly everything else taught at the universities by age 15, when he studied law with John Austin, a utilitarian law professor.[2] James Mill instilled in the young Mill the Benthamite philosophy. Mill was to criticize Bentham in some of his writings, but he never gave up the greatest happiness principle. Mill never held an academic position. When he was 17, his father, who by that time held an important position in the British East India Company, secured a position in the company for his son. John Stuart was to have a career there until the company was nationalized in 1858, by which time he held a

position on a level equal to that of a secretary of state.[3] He never visited India but advised the officers in India by correspondence. The working hours were not excessive, and Mill met with his philosophical friends and did extensive editing and writing while carrying on his duties.

In 1830, when Mill was 24, he met Harriet Taylor (1807–1858), an attractive woman of 22, who was married and the mother of two young children. They developed a "Platonic" relationship, sharing philosophical ideas, especially their shared interest in the liberation of women from subjection in the family and other institutions. Two years after her husband died in 1849, they were married. He called her the co-author of much of his writing. After early retirement from the East India Company, Mill devoted his time to writing, and he served a term in Parliament.[4]

In 1861, when Mill published *Utilitarianism*, he was established as the foremost philosopher and economic theorist of his time and an important political thinker. This reputation was on the basis of his *A System of Logic* (1st edn., 1843), which is not just about logic in the narrow sense but includes a radically empiricist philosophy of language, philosophy of science, and general theory of knowledge, and his *Principles of Political Economy* (1st edn., 1848). He was also a frequent contributor to periodical reviews and had just published *On Liberty* (1859), defending freedom of thought and expression, and liberty of actions and lifestyles that do not harm the legitimate interests of others. Mill's ethical theories were in the context of what John Skorupski calls his "naturalistic" epistemology and metaphysics. Mill thought that "human beings have no supernatural or otherwise non-natural aspect. They belong in the natural order that is studied by science."[5] He was a determinist and found human freedom only in the ability to carry out our choices and desires, including our desires to modify our existing desires if they are not in line with what we want ourselves to be. Mill's epistemology entailed that the only "proof" that can be given for the principle of utility must come from something about scientific human psychology, in his day introspective psychology. It is in that context that one can interpret the "proof" that Mill offers for the principle of utility.

In *Utilitarianism* Mill was to make important and controversial contributions to the Benthamite tradition. Among these, his distinction between pleasures and pains on the basis of "quality" as well as "quantity" was one of the most controversial. Another was his attempt to give a "proof" of the principle of utility. Not so notorious, but very important, was his conception of morality as just one branch of a hedonistic "theory of life," with moral wrongdoing limited to those actions deserving punishment. Still another was his effort to subordinate justice to utility with a theory of rights based on utility. These topics will be discussed in detail later in this chapter.

Utilitarianism at mid-century

Most of the ethical disputes in the nineteenth century were controversies between

defenders of utilitarianism and its critics. In the first half of the century utilitarianism was largely represented in these controversies by William Paley's *The Principles of Moral and Political Philosophy*, published in 1785. Paley (1743–1805) differed from Bentham in that he presented a utilitarianism based on theology. He assumed that the basis of morality is to do the will of God, and he argued that the benevolence of God implies that God's will is the greatest general happiness. The method of coming at the will of God concerning any action, he said, is to determine the tendency of the action to promote or diminish the general happiness. Paley interpreted Scriptures to conform to this view of morality by selecting passages that could be given a utilitarian reading and discounting those that could not.

Paley was attacked by other religious thinkers for his reading of the Bible but also for his moral views. They admitted that God is benevolent, but claimed that God is also just, and that Paley ignored that. More important, they claimed that we have a God-given conscience or moral sense that tells us the difference between right and wrong without having to calculate utility. Paley's book was one of the texts in moral philosophy at Cambridge and the subject of an attack by Adam Sedgwick in *Discourse on the Studies of the University of Cambridge* (1832). Sedgwick's main objection is the immorality of utilitarianism: taking benevolence toward the whole creation as the basis of morality would lead to the destruction of private affections, the ignoring of special duties, the dissolution of marriage, and the end of patriotism.

Mill replied to Sedgwick in a review in 1835[6] and in the review gave his own criticism of Paley. In using Paley as an exemplar of utilitarianism, Mill asserts, Sedgwick has not given the theory its best representative. According to Mill, there are faults in Paley's theory both in its foundations and in its applications. Paley does not take the happiness of mankind to be an end in itself but only a mere index to the will of God, and the motive for morality is not benevolence but only the selfish motive of expectation of rewards and punishment from God. Mill is also critical of Paley's application of the principle of utility. He says that Paley does not use the utilitarian theory as a tool for the criticism of existing morality and legislation, but rather took the doctrines of practical morals that he found current and sought uncritically to provide a utilitarian basis for them.

Mill attacks Sedgwick also for his inability to conceive of progress in morality. He attributes to Sedgwick the theory that we perceive the distinction between right and wrong as we perceive distinctions of color, by a peculiar faculty, the moral sense, or moral instincts. This gives us intuitive principles of morality that are eternal and immutable. In contrast to that, Mill says, is the theory that principles of morality require no other faculties than our intellect and bodily senses. Morality and immorality depend upon the influences of actions upon human happiness, and there can be progress in our knowledge of that. Mill sees this as the chief contrast between utilitarianism and its critics.

Another critic of utilitarianism was William Whewell (1794–1866). J. B. Schneewind,

in *Sidgwick's Ethics and Victorian Moral Philosophy*, says that Whewell's moral philosophy is the most comprehensive system of intuitional ethics produced in Britain during the nineteenth century. A detailed analysis of Whewell's system is not appropriate for a chapter in a work on utilitarianism, but a brief sketch can give an idea of the intuitionistic alternative to utilitarianism. Whewell was a historian of science and believed that progress in science was progress in greater clarification of Ideas in the mind of God. He also thought that there had been progress in morality, and he likewise attributed it to greater clarity in human understanding of Ideas in the mind of God. He believed in a moral faculty, but he thought that it resembled less a passive perception than an active use of Reason. Reasons must be given for claims that actions are obligatory or permitted. Feeling is not moral feeling if it excludes the operation of Reason. The true guide of man is Conscience only so long as the guide of Conscience is Reason. Whewell's system is based on man's use of Reason to control his basic desires, and he classifies these uses into five main virtues: Benevolence, Justice, Truthfulness, Purity, and Order. Purity is the need to control the desire for pleasure, as in the prohibition of adultery. Order is that we must accept the positive statutory laws as the necessary conditions of morality.[7] Whewell began to include lectures on Bentham in his teaching at Cambridge and published these in his *Lectures on the History of Moral Philosophy* (1852).

Mill wrote a review, "Whewell on Moral Philosophy," that year, and Schneewind says that it was this exchange that brought the controversy between utilitarianism and its critics to focus on Bentham's version of utilitarianism rather than Paley's.[8] In his review, Mill gives his chief objection to intuitionistic ethics.

> The contest between the morality which appeals to an external standard [i.e., the standard of utility] and that which grounds itself on internal conviction, is the contest of progressive morality against stationary – of reason and argument against the deification of mere opinion and habit. The doctrine that the existing order of things is the natural order, and that, being natural, all innovation upon it is criminal, is as vicious in morals, as it is now at last admitted to be in physics, and in society and government.[9]

Mill regarded the appeal to intuition or a moral sense as the chief opponent of utilitarian ethics. He also opposed an appeal to what is "natural" as another appeal to prejudice and existing practices. Mill was a reformer. He applied utilitarian thinking to argue for changes in customs and existing institutions. In *On Liberty*, he argued against paternalism. Society should leave people alone if their behavior does not harm other people's legitimate interests. In *The Subjection of Women* (1869), he argued for equality in the marriage relationship, first-class citizenship, and greater economic opportunities for women. In *Considerations on Representative Government* (1861) and other political writings, he expressed the view that representative government best ensured that laws would be made in the interests of the working classes, and he also claimed that electoral

participation made people more active and intelligent than even benevolent authority, cultivating public sympathies and stimulating people to look at questions from impersonal points of view. Any ethics that entrenched existing practices he regarded as vicious. That included appeals to God's will. In his posthumously published *Three Essays on Religion* (1874), he argued against the appeal to Nature and the appeal to the will of God as a basis for ethics.

Mill's *Utilitarianism*

Mill's *Utilitarianism* is a defense of utilitarian ethics. There are five chapters, the first of which, entitled "General Remarks," might be regarded as a preface. The second chapter, "What Utilitarianism Is," presents a succinct formulation of the utilitarian "creed" and then attempts to answer objections to it, objections supposedly based on mistaken interpretations of its meaning. Chapter 3, "Of the Ultimate Sanction of the Principle of Utility," is a discussion of the sources of motivation for conformity to a morality based on the general happiness. Chapter 4 is Mill's presentation regarding "Of What Sort of Proof the Principle of Utility is Susceptible." The final and longest chapter, which Mill had begun writing as a separate essay,[10] is "On the Connexion Between Justice and Utility." This last chapter is in the form of an answer to another objection to utilitarianism, but in this case the objection could better be described as due to an inadequate and incomplete analysis of the idea and sentiment of justice, rather than a mistaken interpretation of utility. Mill's project in the chapter is to show that, when properly understood, justice is consistent with, subordinate to, and an important branch of utility, rather than opposed to it.[11]

In chapter 2, Mill presents a formulation of the utilitarian "creed":

> The creed which accepts as the foundation of morality, Utility, or the Greatest Happiness Principle, holds that actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness. By happiness is intended pleasure, and the absence of pain; by unhappiness, pain, and the privation of pleasure. To give a clear view of the moral standard set up by the theory, much more requires to be said; in particular, what things it includes in the ideas of pain and pleasure; and to what extent this is left an open question. But these supplementary explanations do not affect the theory of life on which this theory of morality is grounded – namely, that pleasure, and freedom from pain, are the only things desirable as ends; and that all desirable things (which are as numerous in the utilitarian as in any other scheme) are desirable either for the pleasure inherent in themselves, or as means to the promotion of pleasure and the prevention of pain.
>
> (2.2/210)[12]

There are several things noteworthy in this formulation. One is the distinction between a theory of morality and a "theory of life" on which this theory of morality is grounded. The theory of life is apparently a general theory of what things are desirable and undesirable as ends, and it is this hedonistic theory of value to which Mill will attempt to persuade the intellect to give assent in his "proof" in chapter 4. The theory of morality that is founded on it is called, in *A System of Logic* and elsewhere, only one branch of the "Art of Life."[13]

Morality is concerned with actions directed toward the end specified as desirable by the theory of life, but not all actions that tend to promote happiness or produce unhappiness are appropriately enforced as morally required or prohibited. Mill makes this explicit in chapter 5. There he says:

> We do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience. This seems the real turning point of the distinction between morality and simple expediency.
>
> (5.14/246)

Here in chapter 2, however, he does not add that restriction, and this has caused some confusion in interpreting Mill's theory of morality.

Another controversy is whether in talking of "actions" Mill has in mind particular actions in unique circumstances or types of actions. Since the mid-twentieth century a distinction has been made between "act utilitarianism" and "rule utilitarianism." According to act utilitarianism, a right action is one that would have as good consequences as any other, given the specific circumstances that exist at the time of the action. According to rule utilitarianism, there are useful rules that, if generally observed, enable people to predict the behavior of others and thus as a community to produce best consequences, and these rules define right and wrong actions. Mill has been interpreted as both an "act utilitarian" and as a "rule utilitarian."

Mill's theory of morality seems to embody both act-utilitarian and rule-utilitarian reasoning. Mill approved a quotation from one of his father's works that in the performance of our duties there are two sets of cases. In one set of cases a direct estimate of the good of the particular act is necessary; and the agent acts immorally without making it. There are other cases of such ordinary and frequent occurrence that they can be distinguished into classes, such as just, beneficent, brave, and so on, to which belong an appropriate rule.[14] It should also be noted that these classes are the names of virtues; so their practice is a practice of the virtues.

In answering the objection that there is not time before acting to calculate and weigh the consequences of a particular action upon the general happiness, Mill replies that

throughout all of human history "mankind have been learning by experience the tendencies of actions . . . 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" (2.24/224). He thinks that the accepted moral rules of his day do admit of improvement, but "to consider the rules of morality as improvable, is one thing; to pass over the intermediate generalizations entirely, and endeavour to test each individual action directly by the first principle, is another" (2.24/224). This passage might be interpreted to imply that moral rules are mere "rules of thumb" by which to most often make correct act-utilitarian decisions. But in the same context, Mill seems to have a stronger conception of at least *moral* rules. He says: "It is truly a whimsical supposition that if mankind were agreed in considering utility to be the test of morality, they would remain without any agreement as to what *is* useful, and would take no measures *for having their notions on the subject taught to the young, and enforced by law and opinion*" (2.24/224, second emphasis added). If there are measures taken for teaching the rules, and rules are to be enforced, there is a social dimension to the rules that makes them more than rules of thumb for the individual utilitarian agent's choice of action case by case. Mill would also not permit there to be moral "free riders." An individual agent is not permitted to benefit from the security of laws and moral rules but not follow them himself. Mill considers, as an example, the murder of someone whose cruel behavior tends to increase human unhappiness. The individual act of murder has consequences that favor it, but

> The counter-consideration, on the principle of utility, is, that unless persons were punished for killing, and taught not to kill; that if it were thought allowable for any one to put to death at pleasure any human being whom he believes that the world would be well rid of, nobody's life would be safe . . . If one person may break through the rule on his or her own judgment, the same liberty cannot be refused to others.[15]

Mill evidently thinks that there is a convergence between act-utilitarian considerations, rule-utilitarian considerations, and the virtues. "The multiplication of happiness is, according to the utilitarian ethics, the object of virtue" (2.19/220). He points out that telling a lie does that much to weaken the useful character trait of honesty and to weaken the transcendently useful rule of veracity. But even this rule admits of exceptions, and he says that there is "no ethical creed which does not temper the rigidity of its laws, by giving a certain latitude, under the moral responsibility of the agent, for accommodation to peculiarities of circumstances" (2.25/225). In "Whewell on Moral Philosophy," however, he had said, "The essential is, that the exception should be itself a general rule; so that, being of definite extent, and not leaving the expediencies to the partial judgment of the agent in the individual case, it may not shake the stability of the wider rule in the cases to which the reason for the exception does not extend."[16]

Mill also recognizes morally meritorious action that goes beyond the "call of duty." He does not want every act to be subject to *moral* evaluation, morally right or wrong. Once basic moral standards have been met, he wants individuals to be free to live their lives without worrying about whether they are doing the act having marginally better or worse consequences than all alternatives. There is room for "supererogation," actions that are morally meritorious but which go beyond the call of duty. In his book *Auguste Comte and Positivism*, Mill says:

> It is not good that persons should be bound, by other people's opinion, to do everything that they would deserve praise for doing. There is a standard of altruism to which all should be required to come up, and a degree beyond it which is not obligatory, but meritorious. It is incumbent on every one to restrain the pursuit of his personal objects within the limits consistent with the essential interests of others . . . If in addition to fulfilling this obligation, persons make the good of others a direct object of disinterested exertions, postponing or sacrificing to it even innocent personal indulgences, they deserve gratitude and honour, and are fit subjects of moral praise . . . but the encouragement should take the form of making self-devotion pleasant, not that of making everything else painful.[17]

Mill's conception of morality is thus complicated. Sometimes morality requires following recognized rules. He wants the practice of some of these rules to be developed into habitual character traits, i.e., virtues. Mill's position is still further complicated by the fact that some rules have correlative rights that are to be respected and that entitle the right holder to make valid claims even if recognition of those claims in a particular case does not maximize utility. This is stated explicitly in a letter to George Grote in 1862, shortly after the publication of *Utilitarianism*. There he says:

> [R]ights and obligations must, as you say, be recognised; and people must, on the one hand, not be required to sacrifice even their own less good to another's greater, where no general rule has given the other the right to the sacrifice; while, when a right *has* been recognised, they must, in most cases, yield to that right even at the sacrifice, in the particular case, of their own greater good to another's less.[18]

Still a further complication arises from the fact that some types of acts that might have best consequences on rare occasions cannot be done but from motives that are incompatible with overall behavior having best consequences. There are two reasons for this. First, states of character are confirmed patterns of behavior that do not permit complete flexibility. The habit of honesty, which is useful as a virtue, cannot be maintained while telling lies on all the occasions when a lie would have marginally better consequences; to be able to lie on every occasion when a lie would maximize utility is incompatible with the most useful degree of habitual honesty. Second, acts presuppose

states of mind that may be states of enjoyment or of wretchedness in themselves. Such mental states are important in a utilitarian calculation, for Mill considers the pleasures of a nobleness of character, of being a person of "feeling and conscience" rather than "selfish and base," to be among the qualitatively higher pleasures (2.6/211). Moreover, these states of mind produce further consequences indirectly, through encouraging the formation of other states of mind: "No person can be a thief or a liar without being much else: and if our moral judgments and feelings with respect to a person convicted of either vice, were grounded solely upon the pernicious tendency of thieving and of lying, they would be partial and incomplete."[19] There is nothing inconsistent about this. The ultimate principle for Mill is promotion of the greatest happiness. Whether to calculate consequences case by case, or to act in accordance with rules, or to respect rights, or to practice justified virtues, is a matter of choosing the appropriate means for the promotion of greatest happiness.

Mill's qualitative hedonism

One of the most controversial innovations that Mill made in utilitarianism was the claim that there are "higher" and "lower" pleasures. Mill and Bentham were both hedonists. They believed that the only things of value and disvalue as ends of actions are pleasures and pains. And they both gave a "mental state" analysis of pleasure and pain. One might be pleased or displeased that some external state of affairs existed or that some external event occurred, but the pleasure or displeasure would be in one's mind, in the consciousness of it.

Bentham analyzed the measure of pleasures and pains along two dimensions. A "lot" of pleasure or pain consists of a certain intensity per moment and a certain duration. He is famous for saying that "quantity of pleasure being equal, push-pin is as good as poetry."[20]

In claiming that there are qualitative differences between pleasures, making some "higher" and some "lower," Mill is responding to the criticism of utilitarianism that in saying that pleasure and pain are the sole criteria of a good life, it is a doctrine worthy of swine (2.3/210). Mill's reply is that if "human beings [were] capable of no pleasures except those of which swine are capable . . . the rule of life that is good enough for the one would be good enough for the other" (2.4/210). But "Human beings have faculties more elevated than the animal appetites, and when once made conscious of them, do not regard anything as happiness which does not include their gratification" (2.4/210–211). The higher faculties that he names are "the intellect . . . the feelings and imagination, and . . . the moral sentiments" (2.4/211). He claims that "It is quite compatible with the principle of utility to recognize the fact, that some *kinds* of pleasure are more desirable and more valuable than others" (2.4/211), not just instrumentally but as immediate pleasurable experiences. His evidence to support this is that

> those who are equally acquainted with, and equally capable of appreciating and enjoying, both [higher and lower pleasures], do give a most marked preference to the manner of existence which employs their higher faculties. Few human creatures would consent to be changed into any of the lower animals, for a promise of the fullest allowance of a beast's pleasures; no intelligent human being would consent to be a fool, no instructed person would be an ignoramus, no person of feeling and conscience would be selfish and base, even though they should be persuaded that the fool, the dunce, or the rascal is better satisfied with his lot than they are with theirs.
>
> (2.6/211)

Mill explains that this is due to "a sense of dignity . . . which is so essential a part of the happiness in those in whom it is strong, that nothing which conflicts with it could be, otherwise than momentarily, an object of desire to them" (2.6/212).

Mill says that the only procedure available for judging which of two pleasures is more worth having is the preference of people who are qualified to judge by having experienced both pleasures. And he claims that there is "no other tribunal" for judgments of quantity, as well, between two pleasures or whether a particular pleasure is worth having at the cost of a particular pain: "the test of quality, and the rule for measuring it against quantity, being the preference felt by those who, in their opportunities of experience, to which must be added their habits of self-consciousness and self-observation, are best furnished with the means of comparison" (2.10/214). The preference of these competent judges does not *constitute* the value of the experiences. It is evidential. It requires practice in self-observation and can be improved by the insights of others. There may be differences in preferences between presumably competent judges, and their preferences may change over time. Majority opinion does not settle the matter, but it is the best evidence for a general statement of the value of different kinds of pleasures.

Critics have asserted that Mill's distinction between kinds of pleasures and pains makes a calculus of pleasure and pain impossible. Mill's introduction of a qualitative dimension does make any calculation more complicated. But Wendy Donner has pointed out that Bentham's quantitative dimensions of pleasures and pains are not as simplistic as they appear. Quantity is a combination of intensity and duration. Bentham assumed that they should be given equal weight, but agents of different character and outlook will differ on how to weigh them. How much intensity in a brief period would outweigh less intensity over a long period of time?[21] Mill's third dimension of higher or lower quality simply makes calculation more complex.

When *Utilitarianism* was first published, Mill was immediately charged with either needlessly complicating hedonism or simply deserting it with his distinction between pleasures on the basis of quality. Either it was quantity under a different name or it was

an appeal to a non-hedonistic value.[22] In these critiques and in so many since then, the basic assumption is made that pleasure is a kind of sensation that feels the same no matter its source; so only the intensity and duration of this one kind of sensation can be grounds for preference. However, this begs the question against Mill. If Mill is correct that there really are introspectively different feelings that are all varieties of pleasure and not of something else, then it is possible for these to be compared on the basis of their felt differences and for some to be preferable to others.

Introspectively Mill appears to be correct that there are qualitatively different pleasures and pains. This is more obvious in the case of pains. If we compare two sensations of pain, such as a stomach ache and a toothache, they are not just in different bodily locations. They feel different as pains. If we compare the pain of grief with the pain of shame, they are different as pains, and both different from a toothache. Why then are they all called pains?

Mill's theory of language accounts for this. With regard to simple sensations, Mill appeals to remembered resemblance to explain the signification of names: "the words *sensation of white* signify, that the sensation which I so denominate resembles other sensations which I remember to have had before, and to have called by that name."[23] Mill does not say that the qualitative resemblance must be identity, so long as it resembles other experiences of that sort more than it resembles an experience of a different sort. In the example, it need not be a sensation of pure white. So long as it resembles other sensations of white more than gray or yellow or some other color sensation, it counts as a sensation of white. By analogy, if qualitatively different pleasures or pains are more like other pleasures or pains than like any other feelings, they are all generically pleasures and pains.

Supposing that Mill is correct that there are qualitatively different pleasures, it does not follow that some are consistently preferred by those adequately acquainted with the different kinds. And even if they are, are the preferred ones correlated with distinctly human faculties? These questions are difficult to answer, for pleasurable and painful experiences are complex, having instrumental as well as intrinsic value and disvalue, and we have second-order attitudes toward our pleasures that may be part of the total experience. This may explain Mill's reference to a sense of dignity that gives superiority to pleasures of the distinctly human faculties. If one is engaging in an activity that one regards as degrading, one may feel a second-order pain as part of the total experience. If one is engaging in an activity in which one feels pride, there may be a second-order pleasure of self-respect. These are not just non-hedonistic values of the experiences. These are hedonistic components of the experiences.

Mill's argument for the superiority of pleasures that involve the higher faculties is not that one would prefer a single experience of a higher pleasure over a greater quantity of a lower pleasure on every occasion of choice, for Mill says that the test of quality, and the rule for measuring it against quantity, is the preference of competent judges, implying

that in at least some cases quantity may outweigh quality. The argument rests on a different question: whether one would be willing to *resign* the higher pleasures for any quantity of the lower; whether one prefers a "manner of existence which employs [the] higher faculties" (2.6/211) or would be willing to "sink into . . . a lower grade of existence" (2.6/212). This is not a lexical ordering of pleasures case by case. And when asked what one would be willing to resign, the question can be reversed. Would one be willing to resign all the pleasures that we share with animals – eating, drinking, sexual gratification, physical exercise, and physical comfort – for any number of pleasures of the intellect?

Mill also says that a happy life consists of "few and transitory pains, many and *various* pleasures" (2.12/215, emphasis added). As long as one does not devote oneself *exclusively* to animal appetites, it would seem that Mill should approve a life of sensory as well as of higher enjoyments.

Mill's "proof" of the principle of utility

In chapter 4 of *Utilitarianism*, Mill discusses "Of What Sort of Proof the Principle of Utility is Susceptible." He says that it is not proof in the ordinary sense, meaning, presumably, that it is not a deductive entailment from premises. It is an argument based on introspective psychology "capable of determining the intellect," and he says that this is "equivalent to proof" (1.5/208).

The evidence on which the argument is based is what people desire as ends. The claim is that when desires are properly analyzed each person desires his or her own happiness, insofar as it is believed to be attainable, and all other things that are desired as ends in themselves, such as possession of money, or power, or even virtue, are desired as "parts" of one's happiness. These are not originally desired as ends, but by association with pleasure and avoidance of pain, individuals become such that they cannot be happy without the possession of these.

Mill recognizes that some actions are done as ends without thought of any pleasure or pain, but he says that these kinds of actions were originally based on desire and are now done from habit. "Will is the child of desire, and passes out of the dominion of its parent only to come under that of habit. That which is the result of habit affords no presumption of being intrinsically good" (4.11/239).

The interpretation and evaluation of Mill's "proof" has been the subject of much discussion. Writing in 1965, J. B. Schneewind said that in the previous fifteen years there had been more essays dealing with the topic of "Mill's Proof" than with any other single topic in the history of ethical thought.[24] The flow of scholarship on this topic has not abated since.

One of the challenges was to Mill's analogy between 'desirable' and 'visible'. Mill says

that the proof that something is visible is that it is seen. The evidence that something is desirable as an end is that it is desired as an end. Mill does not mean by 'desirable' *capable* of being desired, although he has been interpreted that way.[25] He has also been frequently accused of equivocation between "capable of being desired" and "worthy of being desired." The correct interpretation of the analogy between 'visible' and 'desirable' is that both are appeals to evidence – in one case the evidence of the visual sense; in the other case, the evidence of our "desiring faculty."

Mill also generalizes from the fact that each person desires his or her own happiness to the conclusion that the general happiness is what is desirable for the aggregate of all persons. This has been criticized as a fallacy of composition, but in correspondence[26] Mill makes clear that he does not regard the general happiness as anything but a summation of the happiness of the individuals making up the aggregate. If happiness is the *kind* of thing that is desirable, the instances of it in the consciousness of different individuals can be added to constitute what is desirable for an aggregate. Not all present-day philosophers agree that this kind of addition is possible, but it is now generally accepted as Mill's position.

Another source of controversy is Mill's statement that "desiring a thing and finding it pleasant, aversion to it and thinking of it as painful, are . . . two different modes of naming the same psychological fact . . . to desire anything, except in proportion as the idea of it is pleasant, is a physical and metaphysical impossibility" (4.10/237–238). This statement is puzzling to the twenty-first-century reader, but in context Mill is asking the reader to engage in "practised self-consciousness and self-observation" (4.10/237). If the terms were reducible to one another independent of observation, it is hard to see why he would invite one to attempt what appears to be an empirical confirmation. It has also been pointed out that the term 'metaphysical' means approximately 'psychological' to him.[27]

Mill's psychology may be mistaken, but there is now a growing consensus that in his "proof" the author of *A System of Logic* is not committing elementary logical fallacies unworthy of a logician. He is appealing to psychological evidence to move from facts of pleasure and pain and of desires and aversions to judgments of good and bad as ends of actions.

Mill's theory of justice

Mill seems to have originally planned *Utilitarianism* as an essay on the foundations of ethics.[28] But the format is primarily an answer to objections to utilitarian ethics. In chapter 2, he answers various objections: that it is a doctrine worthy of swine, that there is no time prior to action to calculate the consequences of actions, that happiness is unattainable, that utilitarianism renders people cold and unsympathizing, that it is a godless doctrine, and so on. He saves a major objection for chapter 5, the objection that

justice is a moral consideration independent of utility and often in conflict with it. Mill attempts to rebut this with arguments that justice is an important part of utility. He recognizes that the subjective mental feeling, the "sentiment," attached to justice and injustice is different from that which commonly attaches to the general promotion of happiness. Also, except in extreme cases, justice is far more imperative in its demands. He admits that the sentiment is not derived from utility, but in the course of the chapter he argues that what is moral in the sentiment does depend upon utility. And he argues that there is a utilitarian basis for distinguishing justice from other moral obligations and for making the requirements of justice more demanding. If justice is something altogether distinct from utility, which the mind can recognize intuitively, why is there so much controversy over what is just in punishment, in wages, and in taxation? If, on the other hand, justice is subordinate to utility, this is explicable. There will be as much difference of opinion about what is just as about what is useful to society.

Mill examines the etymology and the history of usage of the word. He says that it has an origin connected with conformity to law. It is not attached to all laws, however, but to "such laws as *ought* to exist" (5.12/245), and there may be behavior that is just or unjust where there are no laws that apply. Even here, Mill thinks that the idea of a penal sanction is the generating idea of the notion of justice, but that is true, he says, of all wrongdoing. The distinguishing feature of justice, he claims, is that duties of justice have a correlative *right* in some person or persons. "Justice implies something which it is not only right to do, and wrong not to do, but which some individual person can claim from us as his moral right. No one has a moral right to our generosity or beneficence, because we are not bound to practise those virtues towards any given individual" (5.15/247).

Turning to the feeling which accompanies the idea of justice, Mill says that "the two essential ingredients [of it] are, the desire to punish a person who has done harm, and the . . . belief that there is some definite individual or individuals to whom harm has been done" (5.18/248). He thinks that this desire is derived from two more basic sentiments that "either are or resemble instincts; the impulse of self-defence, and the feeling of sympathy" (5.19/248). He goes on to argue that "It is natural to resent, and to repel or retaliate, any harm done or attempted against ourselves, or against those with whom we sympathize" (5.24/248).

Having analyzed justice as the class of obligations that have correlative rights, Mill gives an analysis of what it is to have a right: "When 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, either by the force of law, or by that of education and opinion" (5.24/250). So far Mill has been analyzing the concepts of justice and of rights independently of the principle of utility. Now he introduces that principle. When asked why society ought to recognize such rights, Mill says that he "can give . . . no other reason than general utility. If that expression does not seem to convey a sufficient feeling of the strength of the obligation" (5.25/250), it is because the feeling includes the "animal element" of self-defense as well as a rational element, and because it is an "extraordinarily important and impressive kind

of utility which is concerned" – that of security. "[S]ecurity no human being can possibly do without; on it we depend for all our immunity from evil, and for the whole value of all and every good, beyond the passing moment" (5.25/250–251).

So far we have Mill's statement that justice is subordinate to utility in the broadest sense. The alternative is that we just know what justice and injustice are independently of utility. Against this idea Mill points to great controversies about what policies, in punishment, wages, and taxation, are just and unjust (5.28–31/251–255). If justice is something that "the mind can recognise by simple introspection . . . it is hard to understand why that internal oracle is so ambiguous" (5.26/251).

Mill still recognizes an important distinction between justice and general utility. He holds

> justice which is grounded on utility to be the chief part, and incomparably the most sacred and binding part, of all morality. Justice is a name for certain classes of moral rules, which concern the essentials of human well-being more nearly, and are therefore of more absolute obligation, than any other rules for the guidance of life . . .
>
> The moral rules which forbid mankind to hurt one another (in which we must never forget to include wrongful interference with each other's freedom) are more vital to human well-being than any maxims, however important, which only point out the best mode of managing some department of human affairs . . . It is their observance which alone preserves peace among human beings . . . a person may possibly not need the benefits of others; but he always needs that they should not do him hurt.
>
> (5.32–33/255–256)

Thus, Mill feels that he has answered the objection that justice is distinct from utility. His claim is that the modes of conduct required by justice can be given a utilitarian justification and, in cases of conflict between competing theories of justice, even require a utilitarian arbitration. And although the sentiment which attaches to instances of justice is different from that which attaches to utility in general, the very existence of that distinct and stronger sentiment has a utilitarian support.

Conclusion

Mill popularized utilitarianism and dignified it by a defense from the greatest British philosopher of his time. He also worked out a complex theory that is not subject to many of the criticisms that have been directed against utilitarianism in its more simplistic forms. His theory is not a maximizing act-utilitarian version of utilitarianism. It has a role for authoritative rules, rights, and virtues. In some details of psychology and argument it may be subject to criticism, but it is a plausible theory that must be taken seriously by ethical philosophers.

Notes

1. J. Mill, *Political Writings*.

2. J. S. Mill, *Autobiography*, *Collected Works*, vol. I, pp. 5–39. See also Packe, *The Life of John Stuart Mill*, pp. 19–50.

3. J. S. Mill, *Autobiography*, *Collected Works*, vol. I, pp. 85–87. See also Packe, *The Life of John Stuart Mill*, p. 80 and pp. 290–291.

4. J. S. Mill, *Autobiography*, *Collected Works*, vol. I, pp. 193–199, pp. 249–261, and pp. 273–285. See also Packe, *The Life of John Stuart Mill*, pp. 123–149, pp. 313–325, and pp. 449–473.

5. Skorupski, "The Place of Utilitarianism in Mill's Philosophy," p. 45.

6. J. S. Mill, "Sedgwick's Discourse," *Collected Works*, vol. X, pp. 31–74.

7. Schneewind, *Sidgwick's Ethics and Victorian Moral Philosophy*, pp. 101–112.

8. Schneewind, *Sidgwick's Ethics and Victorian Moral Philosophy*, p. 131.

9. J. S. Mill, "Whewell on Moral Philosophy," *Collected Works*, vol. X, p. 179.

10. Robson, "Textual Introduction," p. cxxiv.

11. West, *Mill's Utilitarianism*, p. 28 and pp. 89–123.

12. References in the text are to chapter, paragraph, and page of J. S. Mill, *Utilitarianism*, *Collected Works*, vol. X, pp. 203–259. So '2.2/210' refers to chapter 2, paragraph 2, page 210.

13. J. S. Mill, *A System of Logic, Collected Works*, vol. VIII, p. 949.

14. J. Mill, *Analysis*, pp. 312–313.

15. J. S. Mill, "Whewell on Moral Philosophy," *Collected Works*, vol. X, pp. 181–182.

16. J. S. Mill, "Whewell on Moral Philosophy," *Collected Works*, vol. X, p. 183.

17. J. S. Mill, *Auguste Comte and Positivism*, *Collected Works*, vol. X, pp. 337–338.

18. J. S. Mill, "Letter to George Grote," *Collected Works*, vol. XV, p. 762.

19. J. S. Mill, "Remarks on Bentham's Philosophy," *Collected Works*, vol. X, p. 7.

20. J. S. Mill, "Bentham," *Collected Works*, vol. X, p. 113.

21. Donner, "Mill's Theory of Value," pp. 121–122.

22. Grote, *An Examination of the Utilitarian Philosophy*, p. 47.

23. J. S. Mill, *A System of Logic, Collected Works*, vol. VII, p. 136.

24. Schneewind, "Introduction," p. 31.

25. Wall, "Mill on Happiness as an End."

26. J. S. Mill, "Letter to Henry Jones," *Collected Works*, vol. XVI, p. 1414.

27. Mandelbaum, "On Interpreting Mill's *Utilitarianism*," p. 39.

28. Robson, "Textual Introduction," p. cxxiii.

4 Sidgwick and utilitarianism in the late nineteenth century

Roger Crisp

When Henry Sidgwick died, in the last year of the nineteenth century, he was widely thought to be the preeminent moral philosopher of his age. Sidgwick himself was deeply influenced by earlier utilitarian thinkers, including Jeremy Bentham and J. S. Mill, but he took utilitarian ethics to an unrivaled level of sophistication. Discussion around his work, especially his magisterial *The Methods of Ethics* (first published in 1874), continued for about a decade after his death, by which time the focus in utilitarian ethics had shifted onto G. E. Moore. By the middle of the twentieth century, however, there were signs of renewed interest in Sidgwick, and he significantly shaped the moral and political philosophy of the latter half of that century through his influence on John Rawls, Derek Parfit, and others.

Sidgwick's main contribution to philosophy was, without doubt, the *Methods*, which went through several editions. Its publication marked an important move away from the negative tone of much work on utilitarianism in the period, written in response to J. S. Mill's widely read essay *Utilitarianism*, published in book form in 1863 (of course, Mill had his defenders). The last significant edition of the *Methods* was the seventh (1907), and, because that is now the standard edition, it will be my primary text throughout this chapter.[1] But it should be remembered that Sidgwick also wrote several important essays in ethics and a classic history of ethics.[2] I shall begin with an examination of Sidgwick's views on the nature of philosophical ethics, before moving onto his hedonistic theory of well-being and his intuitionist moral epistemology. I shall then discuss his version of utilitarianism and its relation, as Sidgwick saw it, to the common-sense morality of his day. It is misleading to describe Sidgwick as an unqualified utilitarian, since he remained undecided between utilitarianism (or "universalistic hedonism") and egoism (or "individualistic hedonism"). That topic – his so-called "dualism of practical reason" – will be discussed in my final section.

The nature and methods of ethics

Philosophical ethics, Sidgwick tells us, like science aims to be "systematic and precise" (1.1.1.2/1). He later says that the assumption that moral rules should be precise "naturally belongs to the ordinary or jural view of Ethics as concerned with a moral code" (3.2.3.1/228), and he provides an argument for this view based on an apt analogy with law. If a law were vague, we would think it to that extent unreasonable: anyone subject to a legal obligation ought to be in a position to know what it is. Similarly, a moral philosophy which left it unclear on some occasion exactly what a person's obligations

were would, to that extent, have failed.

A good deal of law is indeed precise. The UK Representation of the People Act 1969, for example, leaves no doubt about when a person becomes eligible to vote in a parliamentary election: on their eighteenth birthday. But some law is less precise. Consider the definition of obscenity in the Obscene Publications Act 1959, still in force in the UK: what tends to "corrupt and deprave." If I have written some potentially obscene article, and am considering its publication in the UK, I have to rely on my judgment about the likely effects of its publication, and whether they might be described as depravation or corruption. A law is not a failure if it is *reasonably* clear, and relies only to a *reasonable* degree on individual judgment. Exactly what counts as reasonable or not in ethics is a highly important question, and it may be said that Sidgwick's standard of reasonableness is too high and that this leads him to a somewhat uncharitable view of the resources of non-utilitarian ethical theories.

The scope of philosophical ethics, Sidgwick says, is the "methods" of ethics, where a method is "any rational procedure by which we determine what individual human beings 'ought' – or what it is 'right' for them – to do, or to seek to realize by voluntary action" (1.1.1.1/1). Such a "procedure" need not be a process: immediate insight into the rightness of certain actions is a method (1.1.2.3/4). Nor is Sidgwick to be understood as suggesting that only actions matter in ethics, and not, say, the feelings or the characters of agents. Indeed he elsewhere allows that the common-sense conception of virtue includes the emotions (3.2.2.2/222–223) and that ethics should construct ideals of character (3.14.1.3/393). But discussion of such topics is significant, for Sidgwick, only insofar as it is related to the primary question of ethics – how we should act. Like Aristotle, Sidgwick sees philosophical ethics as essentially practical.[3]

The definition of method above leaves it open whether Sidgwick is including ethical theories, which advocate certain basic normative principles, as among the "rational procedures" he has in mind. It appears not. Consider the view that God has implanted in us knowledge of certain apparently non-utilitarian common-sense rules, such as the rule that we should keep promises, because these rules are the best way to promote the utilitarian end of general happiness. According to Sidgwick, this view constitutes a *rejection* of the *method* of utilitarianism, though not of the utilitarian *principle* (1.6.3.4/85). But since philosophical ethics is an inquiry into what grounds or justifies our actions and any decision-procedure we adopt, we might wonder why Sidgwick emphasizes methods rather than ultimate principles. Ethics, as Sidgwick points out, is "sometimes considered as an investigation of the true . . . rational precepts of Conduct" (1.1.2.1/2–3), and he himself implies that we are interested in the principles that determine which conduct is ultimately reasonable (1.1.3.5/5–6). Sidgwick's book should perhaps have been titled *The Ultimate Principles of Ethics*, those principles each being a different statement of our ultimate reasons for action.

According to Sidgwick, we continually inquire into what is ultimately reasonable

because different and incompatible principles are present in common practical reasoning (1.1.3.5/6), and an answer given in terms of one of them will appear suspect from the point of view of the others. Which principles does Sidgwick have in mind? There are three.

The first set of principles are those that constitute the morality of common sense (1.1.4.3–4/7–8): the rules of prudence, justice, veracity, and so on. Sidgwick suggests that common sense ordinarily sees these rules as binding in any particular case independently of the consequences of the action in question or its alternatives, and they provide the basis for what he calls "dogmatic" *intuitionism*. Sidgwick then notes (1.1.1.5–6/8) that many *utilitarian* thinkers see these common-sense rules as mere means to the general happiness of humanity or sentient beings, and contrasts this view with the *egoistic* principle of prudence which rests on the postulation of the individual's happiness as an end (this is "ethical" in the broad sense that it states a principle concerning ultimate reasons for action).

Sidgwick does not seriously consider versions of egoism and utilitarianism which propose the pursuit of non-hedonistic, non-moral goods, such as knowledge or accomplishment. And he categorizes versions which advocate the pursuit of human excellence as special types of intuitionism, because many past thinkers have seen virtue as a preeminent component of perfection, and the promotion of virtue is central to several such forms of intuitionism. But even if the views are extensionally equivalent, they should not be identified, and it may be that Sidgwick's failure to see this is connected with his excessive stress on methods instead of principles. According to intuitionism, the reason I should keep my promise, for example, is that promises should be kept. According to perfectionism, however, the ultimate reason here consists in the promotion of perfection.

Further, it is tempting to think that Sidgwick is allowing his own substantive views to "filter" the deliverances of common sense at an early stage in the argument: a footnote to the sentence in which he says he will treat perfectionism as a form of intuitionism refers us to the argument in 3.14 against non-hedonistic conceptions of the good. Sidgwick claims he is taking his three main methods from an impartial review of common sense, but perhaps the most plausible explanation for his discussing egoism, common-sense morality or intuitionism, and utilitarianism is that these are the three views *he* finds most plausible (1.1.5.2/14). But this does not undermine his project. Moral theories can plausibly be distinguished as broadly consequentialist on the one hand, and non-consequentialist or deontological on the other. And the main opposition to moralism is, of course, rational egoism, which, though it is not currently much discussed in contemporary ethics, remains one of the most powerful and attractive normative theories. Indeed, Sidgwick could plausibly have started with these three theories, and gone on to explain how different accounts of the good will give rise to different forms of egoism and utilitarianism in particular. His arguments for his basic triad are problematic, but it stands independently of them.

Hedonism

Sidgwick's tripartite division of ethical theories explains why the *Methods* is divided into four separate books. The first is introductory, the second concerns individualistic or egoistic hedonism, the third intuitionism (the non-consequentialist ethical theory developed on the basis of common-sense morality), and the fourth universalistic hedonism or utilitarianism. But his discussions of each theory, and the issues they raise, are not restricted to their respective books.

Sidgwick's main discussions of hedonism are in connection with egoism. According to *rational egoism*, the only reason I have to act in any way depends on the extent to which that action furthers my own good, well-being, or welfare. But, to become practical, any such version of egoism must be combined with an account of what that good consists in. According to what we might call *welfare hedonism*, the only positive constituent of well-being is pleasure or enjoyment, and the only negative constituent is pain or suffering, a natural implication of the view being, of course, that the best life for any individual consists in that life with the greatest balance of pleasure over pain. Welfare hedonism combined with rational egoism will give us the view to which most of book 2 of the *Methods* is dedicated: *egoistic hedonism*. Welfare hedonism is consistent with the view that there are goods other than pleasure, and "bads" other than pain, as long as these goods and bads are not seen as constituents of well-being. So I might believe that, though a person's life can be improved only through an increase in the overall balance of pleasure over pain within that life, a *world* can be improved through, say, an increase in its beauty (beauty being understood as a good in itself). According to *global hedonism*, there are no such non-hedonistic goods or bads. Sidgwick accepted both welfare and global forms of hedonism: the only ultimate good is desirable consciousness (3.14.3.2/397). Sidgwick concentrates in particular on welfare hedonism, though it is worth noting that the arguments for and against welfare hedonism often carry across directly to the global form of the theory.

Sidgwick's arguments for hedonism need to be understood in the context of his overall epistemological position, which will be discussed further in the "Intuitionism" section below. One major component of that position is his philosophical intuitionism, according to which certain propositions are "self-evident" and a person who properly and reflectively grasps them can be justified in believing them on the basis of that grasp. Like many hedonists, Sidgwick tends not to begin with positive arguments for the hedonistic position before moving on to consider alternative positions and objections. Rather he outlines his arguments in response to these positions and objections.

In his important chapter "Ultimate Good" (3.14), Sidgwick faces the question whether the ultimate good is desirable consciousness, with virtuous action as one component, and perhaps other elements also, including "physical action, nutrition, and repose." Having dismissed physical processes and evolutionary fitness as candidates for the ultimate good, he firmly states his view that virtue is valuable only because of the pleasure it produces

for the agent and others. Many, however, will not agree that pleasantness is the only property relevant even to the evaluation of feeling. Consider, for example, the sense of awe experienced by a woman gazing at her newborn baby. Indeed the idea that there are non-hedonic welfare values – whether consisting in states of consciousness or not – is the central objection to hedonism. Consider, for example, the idealist F. H. Bradley's suggestion that those not in the grip of a philosophical theory will be inclined to think that "there are things 'we should choose even if no pleasure came from them'."[4]

Recognizing this, Sidgwick appeals to his reader to reflect upon her own considered intuitions, and also to consider the judgments of humanity as a whole (3.14.5.1–3/400–402). Here we see, as well as Sidgwick's philosophical intuitionism, his Aristotelian commitment to testing one's intuitions "dialectically" against the views of others. Sidgwick notes that the non-hedonic or "ideal" goods do produce pleasure in several ways, and that common sense approves of them roughly in proportion to the degree of such productiveness. But for a clear denial of Sidgwick's position on beauty, one not clearly inconsistent with common sense, consider for example G. E. Moore's case of the beautiful universe which Moore believes has value even if it is never seen by anyone.[5] Sidgwick accepts that knowledge is harder for the hedonist to deal with, but notes for example that common sense is especially impressed by knowledge that bears fruit.

Sidgwick might also appeal to his later "debunking" account of common-sense morality, suggesting that the principles concerning evil, nobility, authenticity, and so on, on which many such objections rest, are themselves grounded on the hedonistic value which their adoption promises. Sidgwick must and indeed does accept that common sense is not, in the end, entirely hedonistic. But he adduces four considerations which are intended to explain this and hence take the sting out of this aspect of common sense (3.14.5.4–10/402–406):

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

have hedonistically valuable consequences also).

Sidgwick's hedonism came under vigorous attack from other moral philosophers of his day, as had that of Mill before him. Sidgwick deals effectively with several contemporary criticisms at 2.3.1.2–2.3.2.2/132–137, including T. H. Green's claim that "pleasure as feeling, in distinction from its conditions that are not feelings, cannot be conceived," and that pleasures cannot be added since they occur in series and not at the same time.[6] The real problem with hedonism, however, is the possibility of non-hedonic or ideal goods, and, as we shall see in the next section, Sidgwick's own epistemology appears to commit him, in the case of all objections relying on the notion of ideal goods, to suspension of judgment, since many reflective thinkers believe in ideal goods and there is no plausible account available of how such thinkers could be mistaken.

Intuitionism

Sidgwick believes that we have ethical knowledge. How does he think we acquire it? In the nineteenth century, philosophers who believed in a human intuitive capacity were often seen in opposition to so-called *inductivists*. Having presented a broad conception of intuition as "immediate judgment as to what ought to be done or aimed at," Sidgwick is careful to situate his own intuitionism (as a purely *epistemological* view, not the ethical version based on common-sense morality) in the context of contemporary debate (1.8.1.2/97–98). Sidgwick rightly notes that the parties in the debate were commonly talking at cross purposes, since each was claiming to know different things. Inductivists claimed inductive knowledge of the pleasantness of certain actions, whereas intuitionists focused on the rightness (or wrongness) of those actions. Any ultimate evaluative or normative view, Sidgwick thought, cannot itself be known inductively. It must be either grasped intuitively, or inferred from other premises at least one of which must include a basic intuition.

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

Sidgwick describes three main categories of intuitionism. The first is *perceptual intuitionism*, according to which conscience tells us what to do in each individual case. He goes on to point out that few will accept such a position, since most people find their own particular intuitions open to doubt, non-comprehensive, inconsistent over time, and

indeed often in conflict with those of others.

Sidgwick distinguishes the next ethical "phase" of intuitionism through reference not to some further moral faculty, but to the object of the faculty in question – namely moral rules. On this view, I might "see" that, for example, promise-breaking is wrong in itself. This is *dogmatic intuitionism*, according to which general moral rules are implicit in common-sense moral thought, and the task of the philosopher is to elucidate and systematize them as far as possible.

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

A defender of dogmatic intuitionism might object that her theory does indeed provide a "deeper explanation" of why certain conduct is right (1.8.4.1/102). Promise-breaking, for example, can be seen to be wrong in itself, or wrong because unjust; helping others is good itself, or because it is benevolent; and so on. For Sidgwick, however, such a theory is still unacceptably unsystematic, and his long and insightful discussion of common-sense morality in book 3 is intended to demonstrate that this is true across the board.

Sidgwick's preferred version of the view is the third: philosophical intuitionism. On this view, any moral knowledge must consist in or at least rest on self-evident, foundational, non-inferential beliefs – that is, intuitions. Sidgwick suggests four conditions that any intuition would have to meet to achieve "the highest degree of certainty" (3.11.2.1/338): clarity and precision; reflectiveness; consistency; and what we might call "non-dissensus." According to that final condition:

> if I find any of my judgments, intuitive or inferential, in direct conflict with a judgment of some other mind, there must be error somewhere: and if I have no more reason to suspect error in the other mind than in my own, reflective comparison between the two judgments necessarily reduces me temporarily to a state of neutrality.

As with hedonism, Sidgwick did not confront the implications of the non-dissensus condition for his own normative ethics in general. He does in 3.13.4 seek to show that his own allegedly self-evident axioms find some support in the work of Samuel Clarke and Kant, and of course the utilitarians. But he fails to reflect upon the fact that many thinkers will disagree with each of his alleged axioms. Such disagreement, indeed, is implicit within his "dualism of practical reason" itself: egoists will disagree with the

principle of benevolence, while utilitarians will rebut the principle of prudence.

It remains a mystery why Sidgwick appears not to have recognized the skeptical implications of his non-dissensus condition. One possible explanation is that he might have believed that one of its implications was a form of skepticism which would have paralyzed his philosophy. But this is not the case: he could have said everything he wanted to say, but on the understanding that what we were getting were Sidgwick's own reports from the fronts of reflection about how things appeared to him. Still, it must be admitted that the non-dissensus condition does "open a door to universal scepticism" (see "Concluding Chapter" (CC) 5.3/509), and Sidgwick clearly found unwelcome the thought of that door's opening to him.

Utilitarianism

As now, in Sidgwick's day the name 'utilitarianism' was used to refer to several quite different positions. So Sidgwick begins his fourth book with a chapter explaining how he understands the view, and how it differs from various other positions. He defines utilitarianism or "Universalistic Hedonism" (UH) as the claim that "the conduct which . . . is objectively right, is that which will produce the greatest amount of happiness on the whole" (4.1.1.2/411).

Consider the following case:

> *The Rash Doctor*. You are suffering from some painful medical condition, for which two drugs are available. Drug A might cure you completely, and there is a 1-percent probability of its doing so. But there is a 99-percent probability of its killing you. There is a 100-percent probability that drug B will almost cure you (though it will leave you with a very slight twinge of pain, once every year or so). Your doctor, in full awareness of these facts, prescribes drug A, and it cures you completely.

On Sidgwick's view, not only is the rash doctor's action objectively right, but the prescription of drug B would have been objectively wrong. This may seem highly counter-intuitive, but we should remember that Sidgwick keeps the notions of wrongness and blameworthiness distinct. Blaming and praising, as activities, are themselves to be regulated by UH, and it is clear that blaming doctors who take such risks with their patients will be required (4.3.2.3/428). In other words, the ideal agent, according to UH, will be the one who always performs actions that maximize happiness overall. The question of which strategy will bring each of us closest to that ideal is, of course, a difficult matter, and much of book 4 is concerned with it.

The second section of 4.1 discusses the scope of utilitarianism, and its relation to equality, foreshadowing debates that became central in the last few decades of the twentieth century concerning the nature of egalitarianism and the moral status of non-

human animals and future generations. Sidgwick begins by repeating his assumption that pleasures and pains must be commensurable, allowing for a certain degree of vagueness: "each may be at least roughly weighed in ideal scales against any other" (4.1.2.1/413). He then claims, plausibly enough, that it would be "arbitrary and unreasonable" not to include the happiness of non-humans within the scope of utilitarianism, noting that the difficulty of assessing non-human experience will arise for any ethical view which does not ignore it.

At this point, Sidgwick makes a highly significant breakthrough in our understanding of utilitarianism. Having noted that UH must be temporally neutral, since the value of happiness does not depend on when it exists, Sidgwick asks what the implications of utilitarianism are regarding population size. It might be thought (and, Sidgwick says, Malthusians indeed have thought) that we should aim to maximize average happiness. But this is to ignore the fact that adding extra members to a population, if their lives are of positive value, may increase the total level of happiness even if the average happiness in the new population is lower than that in the original.

Sidgwick fails to mention that a readiness to sacrifice quality or level of well-being for a greater overall total brought about by increased temporal quantity results in implications such as Derek Parfit's *Repugnant Conclusion*:

> For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better even though its members have lives that are barely worth living.[7]

One way to avoid the Repugnant Conclusion is through the notion of discontinuity of value, according to which losses in quality take on an additional significance that prevents them from being outweighed by increases in the temporal quantity of well-being. But Sidgwick denies discontinuity (2.2.1.1/123–124n1) and it is tempting to think that he would have seen the Conclusion as just another implication of utilitarianism counter to the morality of common sense but not to properly grounded ethical theory.

The chapter ends with an interesting if puzzling discussion of how utilitarians might choose between different distributions of the same amount of happiness. As Sidgwick notes, this question is not purely academic, since the vagueness of actual hedonistic calculations will often result in our being unable to see any quantitative difference between two or more distributions, even if in fact there is one. As Sidgwick also points out, utilitarianism will be indifferent between any set of distributions all of which maximize happiness overall. And so:

> we have to supplement the principle of seeking the greatest happiness on the whole by some principle of Just or Right distribution of this happiness. The principle which most Utilitarians have either tacitly or expressly adopted is that of pure equality – as

> given in Bentham's formula, "everybody to count for one, and nobody for more than one." And this principle seems the only one which does not need a special justification; for . . . it must be reasonable to treat any one man in the same way as any other, if there be no reason apparent for treating him differently.
>
> (4.1.2.6/416–417)

This brief but fascinating passage raises several questions. First, why does Sidgwick feel entitled to interpret Bentham as an egalitarian? Second, why does he see a need for a supplementary principle at all? It is not as if the utilitarian answer in such cases is unclear: any distribution that maximizes is acceptable. Third, if this principle of equality is itself independent of the utilitarian principle, why does it come to play only as a tie-breaker? Why can it not, like egoistic hedonism, be found in practical conflict with the utilitarian principle? Consider the following pairs of possible distributions:

Case I

(a)	
Person A	*Person B*
50	50
(b)	
Person A	*Person B*
1	99

Case II

(a)	
Person A	*Person B*
50	50
(b)	
Person A	*Person B*

On Sidgwick's view, the equality principle requires us to choose (a) over (b) in case I. But it has no force in case II, where we have to choose (b) over (a) on utilitarian grounds. But if equality has no weight against even the smallest amount of utility, one cannot help but wonder why it should have any weight in choices between distributions of equal utilitarian value.

How exactly does Sidgwick ground utilitarianism on intuition? He claims that the "axiom of Rational Benevolence is . . . required as a rational basis for the Utilitarian system" (3.13.5.1/387). That axiom is:

> *RB:* Each one is morally bound to regard the good of any other individual as much as his own, except insofar as he judges it to be less, when impartially viewed, or less certainly knowable or attainable by him.

If this principle is understood monistically, and not as one among several others in some version of dogmatic intuitionist pluralism, it might appear to provide not just a basis for utilitarianism, but a statement of it. Sidgwick, however, as we have seen, prefers to define utilitarianism as a version of hedonism, so for *RB* to constitute a statement of utilitarianism would require substituting 'happiness' for 'good'. And it is that substitution which is required for *RB* to become sufficiently concrete for it to constitute a practical method of ethics (3.13.5.5/388; 4.2.1.4/421n1). How will that substitution be carried out? It is done in the passage advocating hedonism we have already discussed, in which Sidgwick appeals to the intuitions of his reader and to the judgments of humanity in general. In other words, Sidgwick's utilitarianism is established on the basis of combining *RB* with hedonism, each of which is justified via philosophical intuitionism and Aristotelian dialectic. The dialectical support for utilitarianism comes from its relationship to common-sense morality, as we shall now see.

Utilitarianism and common-sense morality

Sidgwick is well aware that his philosophical intuitionist argument will not always persuade those who see views such as egoism or dogmatic intuitionism as self-evident. If a utilitarian is to prove her principle (inferentially) to someone who accepts certain other principles, "the process must be one which establishes a conclusion actually *superior* in validity to the premises from which it starts." I take it that by 'validity', here, Sidgwick means something like "credibility." So, when there is a conflict between utilitarianism and the principles of common-sense morality or the requirements of rational egoism, which the dogmatic intuitionist and the egoist respectively took to be self-evident, the (now ex-)

intuitionist and (ex-) egoist will find utilitarianism more persuasive: "Utilitarianism . . . must be accepted as overruling Intuitionism and Egoism."

In other words, intuitionist or egoist principles have to be given some epistemic weight in the argument, or the proof will not be addressed to intuitionists or egoists. So Sidgwick proposes "a line of argument which on the one hand allows the validity, to a certain extent, of the maxims already accepted, and on the other hand shows them to be not absolutely valid, but needing to be controlled and completed by some more comprehensive principle" (4.2.2/420). For example, an egoist who claims that his own pleasure is good from the point of view of the universe may be persuaded that his happiness is no more important a part of the good overall than that of anyone else.

These coherentist arguments have weight in themselves, as part of the process of Aristotelian dialectic which Sidgwick runs – somewhat unstably – alongside the methodology of philosophical intuitionism. In other words, Sidgwick is granting "validity, to a certain extent" to common-sense morality because he does indeed believe it to have credibility in itself (4.3.1.3/425; see 3.13.1.2/373).

As far as dogmatic intuitionism is concerned, the utilitarian has first to demonstrate that the common-sense principles of veracity, justice, and so on have merely "a dependent and subordinate validity," in that *some* higher principle is required to explain exceptions as well as resolve indeterminacy and conflict (4.2.4/421–422). This "negative" aspect of the argument Sidgwick claims he has "sufficiently developed" in book 3, while the positive stage will be carried out in book 4, where utilitarianism will be shown to be the higher principle in question. The essence of Sidgwick's view on common-sense morality, then, is similar to that of J. S. Mill: it consists in a set of "secondary principles" the ultimate normative justification for which is their promotion of the greatest overall balance of pleasure over pain.

A central problem for Sidgwick is that the dogmatic intuitionist can accept nearly everything Sidgwick says about the general usefulness of common-sense rules and dispositions, but deny the move to utilitarianism as the correct account of the criterion of rightness. It could be argued that, just as Sidgwick took utilitarianism further than any previous thinker, so W. D. Ross, in the twentieth century, did the same for dogmatic intuitionism. Consider the following:

> If, so far as I can see, I could bring equal amounts of good into being by fulfilling my promise and by helping some one to whom I had made no promise, I should not hesitate to regard the former as my duty [and] normally promise-keeping . . . should come before benevolence.[8]

Once one admits that any reasonable moral theory – including utilitarianism – is going to require some not inconsiderable capacity for judgment in individual cases for its application, there seems no good reason for Sidgwick to dismiss a properly reflective

intuitionism such as Ross's. Sidgwick does not, then, have a watertight argument for his utilitarianism. But the same problem arises for any moral theory (including dogmatic intuitionism, of course), and Sidgwick is surely right to think that at least some of his readers may be prompted to make the move to utilitarianism after reflection on the utility-value of common-sense morality.

Sidgwick is interested not only in the question of whether utilitarianism can be proved. As we have seen, he sees philosophical ethics as importantly practical, and in the case of any particular principle he will ask what its practical implications are for making real-life decisions. The obvious utilitarian method is, of course, a practical hedonism based on experience of pleasure and pain: empirical hedonism (4.4.1). In book 2, Sidgwick brilliantly teased out the difficulties in applying this method directly in the case of a single individual. But, he suggests, despite the fact that interpersonal comparisons of utility will lead to yet another level of complexity, it might be thought that the arguments he has offered for utilitarianism's underlying common-sense morality enable us to see the principles of common-sense morality as "middle axioms" of the utilitarian method, lying between the more fundamental principle of utilitarianism itself and individual decisions about what to do in particular circumstances (4.4.1.2/461).

Common-sense morality, however, is not perfectly in alignment with utilitarianism at every point. It is clear, for example, that common-sense predictions of hedonic consequences have frequently been deeply mistaken, because of limited sympathy and knowledge, and distortions resulting from social hierarchies and false religions. And of course predicting such consequences will also be difficult even for the committed utilitarian. Sidgwick believes that defining precisely the moral code recommended by utilitarianism will face problems of the same magnitude as he found with the application of empirical hedonism in book 2 (4.4.2). The nature of human beings themselves varies greatly over time and space, which makes it impossible to construct a universal ideal. If we restrict ourselves to our own time and country, Sidgwick suggests, we face a dilemma: if we take our fellow-citizens as they are, we cannot see them as subjects for a new moral code; while if we abstract away, it is not clear why we should want to design a community for such imaginary beings, unless we assume – quite implausibly – that any code we design will be immediately and appropriately adopted by all. It has to be said that, though Sidgwick is of course right about the difficulties of predicting the future, he makes rather heavy weather of such so-called dilemmas. It is clear that we have to take human beings as they are, while recognizing that human nature includes a capacity to change to some limited degree (as he himself sees: 4.4.3.6/474; 4.5.1.3/475–476; 4.5.1.3/477).

Nevertheless, there is much plausibility in Sidgwick's objections against Herbert Spencer's quasi-Kantian suggestion that moralists should base their practical ethics for this world on the rules of some perfect form of society (4.4.2.4/470–471).[9] In the second half of the nineteenth century, it was quite common for those influenced by Darwinian theories of evolution to see the end of morality as the preservation of the social organism

rather than its happiness (4.4.3). Like Sidgwick, Spencer saw ethics as importantly analogous to science, but the points of analogy he sought to establish were quite different. On Spencer's view, the aim of science is to ascertain general laws of nature through abstraction. So, for example, scientists seek to arrive at fundamental truths of mechanics by ignoring real-life complications arising from friction, plasticity, and so on. In the same way, he believed, moral philosophers should aim to describe the "absolute ethics" of a perfect society, and then use that to create the "relative ethics" that should guide everyday decision-making in the world as it is. As Sidgwick points out, however, we cannot predict the natures and relations of the individuals in a perfect society sufficiently precisely, and even if we could it does not follow that the rules of such a society will provide guidance on rules for us (consider, for example, the need we have for rules concerning punishment).

Again, Sidgwick's objections to Leslie Stephen's version of this view are right on target (4.4.3.1–4/471–473): there is no reason to think that preservation maximizes happiness, since many pleasures and pains have nothing to do with preservation; and, given the imperfect state of sociology, the difficulties in working out the best means to preservation are no less than those in working out those to maximum happiness.[10]

The dualism of practical reason

In his "Concluding Chapter," Sidgwick draws together the strands of his argument with a view to making a final decision on the relation of the three methods with which he began the *Methods*: egoism, intuitionism, and utilitarianism. He begins by reminding us of the (philosophical) intuitionist basis of utilitarianism, and the lack of it for (dogmatic) intuitionism, and that the virtues of dogmatic intuitionism can be seen – partly through reflection on the comparative history and origins of morality – as grounded in impartial benevolence or prudence (CC.1.1/496–497). The question, then, is that of the relation between egoistic and universalistic hedonism, and the challenge for anyone who wishes to argue for the rationality of morality is to demonstrate a harmony between those two views (CC.1.2/497–498).

What does Sidgwick mean by "harmony"? First, we need to be clearer about exactly what he means by egoism and utilitarianism. We can create canonical forms of each by incorporating hedonism into the statements of the principles of prudence and benevolence in 3.13:

> *Egoism (E):* One ought to aim at one's happiness on the whole.
> *Utilitarianism 1 (U1):* Each one is morally bound to regard the happiness of any other individual as much as his own, except insofar as he judges it to be less, when impartially viewed, or less certainly knowable or attainable by him.

To make the contrast between the two views easier to see, we might rephrase this as:

Utilitarianism 2 (U2): One ought to aim at the greatest happiness on the whole.

Sidgwick must have been tempted to reject egoism. He certainly finds it repellent (3.1.1.1/199–200). Egoism sees the rules of common-sense morality as mere means to the end of individual happiness, to be ignored when self-interest requires it, and the view also fails to deliver clear practical guidance: "A dubious guidance to an ignoble end appears to be all that the calculus of Egoistic Hedonism has to offer." Nevertheless, in the end he finds himself unable to reject the position outright:

> 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.
>
> (CC.1.2/498)

How, then, might we reconcile egoism and utilitarianism? It seems clear that Sidgwick is not taking either of these views as attempts at stating the complete truth about normative reasons. If they were, they would be straightforwardly contradictory, since *E* would imply that each of us has reason to promote her own happiness alone, while *U2* would imply that each of us has reason only to promote the happiness of all. Nor is each view to be understood as expressing merely a *pro tanto* reason for action, such that each principle could be balanced against the other in particular cases – in the same sort of way as principles are balanced in Ross's intuitionism. Rather, each is making a claim about what it is ultimately reasonable overall to do. The claim that it is ultimately reasonable to promote one's own happiness is consistent with the claim that it is ultimately reasonable to promote the greatest happiness overall, if promoting one's own happiness is in fact the same as – that is, extensionally equivalent to – promoting the greatest happiness overall. If they are the same, the following claims might be true:

> *E*:* It is ultimately reasonable to promote one's own happiness – that is, to promote the happiness of all.
> *U2*:* It is ultimately reasonable to promote the happiness of all – that is, to promote one's own happiness.

If they are not, we will face a contradiction:

> *E**:* It is ultimately reasonable to promote one's own happiness – that is, not to promote the happiness of all.
> *U2**:* It is ultimately reasonable to promote the happiness of all – that is, not to

promote one's own happiness.

Because complete coincidence is required for such harmony, it is not enough to point out that both *E* and *U2* recommend *general* adherence to the rules of common-sense morality (CC.2). Sidgwick goes on to discuss the claim by some utilitarians, including Mill, that this coincidence is ensured by the priority of sympathy as a component of human happiness (CC.3). Sidgwick reiterates his distinction between sympathy's role in producing pleasures and pains and its role in causing an impulse to action. As he points out, for sympathy to guarantee the coincidence of *E* and *U2*, it must not merely motivate altruistic action but provide maximal happiness for the agent.

Sidgwick recognizes that sympathy can be a source of happiness, and that such sympathy tends to play a role, in the mind of a utilitarian, in the "moral feelings" that concern social conduct. This enables the utilitarian to avoid the objection (often made against Kantian theories in particular) that her theory requires her to sacrifice herself to an "impersonal law" rather than for others she cares about. In an unusually moving passage (Sidgwick himself was considered a person of the highest moral integrity), he claims also that most people's happiness would in fact be promoted were they to cultivate a greater degree of sympathy:

> [T]he selfish man misses the sense of elevation and enlargement given by wide interests; he misses the more secure and serene satisfaction that attends continually on activities directed towards ends more stable in prospect than an individual's happiness can be; he misses the peculiar rich sweetness, depending upon a sort of complex reverberation of sympathy, which is always found in services rendered to those whom we love and who are grateful. He is made to feel in a thousand various ways, according to the degree of refinement which his nature has attained, the discord between the rhythms of his own life and of that larger life of which his own is but an insignificant fraction.
>
> (CC.3.2/501)

But even this is insufficient to provide the complete coincidence required between *E* and *U2* (CC.3.3/501–502; see 2.5.4). A sacrifice of one's life, for example, for the general good could not plausibly be said to advance one's happiness, and the fact that our most intense sympathy is for those close to us increases the motivational opposition to impartial utilitarian duty (nor should we think that attempts to increase the impartiality of our sympathy would be themselves recommended by utilitarianism). The same is true in less unusual cases. Alleviating the suffering of others, for example, will be required by utilitarianism, but sympathy here will be if anything a source of pain rather than pleasure to the agent, and, though it may be counterbalanced by the pleasures of benevolence and so on, an alternative life would often be hedonistically more valuable for the agent.

Sidgwick then moves to another argument put forward by utilitarians of his day: that

utilitarianism is the law of God, to be enforced through a system of divine reward and punishment that will underpin the coincidence of *E* and *U2* (CC.4). This raises the question of what justifies such beliefs. Sidgwick sees the issue of revelation as beyond his remit, though he cannot resist pointing out that most arguments from revelation have been to non-utilitarian conclusions. Sidgwick rejects the view that moral rules should be seen merely as the commands of a divine lawgiver on the ground that God himself is understood as a moral agent bound by the rules, though he is prepared to entertain the view that through intuition we may learn that God commands us to obey certain moral rules grounded independently of his commands or to pursue the same end as he himself pursues – that is, universal happiness.

Can belief in the existence of such a God be justified on purely ethical – that is, rational – grounds alone (CC.5)? Sidgwick has to confess, with some regret, that he does not see it as self-evident that performance of duty will be rewarded and violation punished, whether by God or in any other way. Thus he feels forced

> to admit an ultimate and fundamental contradiction in our apparent intuitions of what is Reasonable in conduct; and from this admission it would seem to follow that the apparently intuitive operation of the Practical Reason, manifested in these contradictory judgments, is after all illusory.
>
> (CC.5.1/508)

In other words, if egoism and utilitarianism, when construed in the light of the facts, contradict one another, neither of them can be said to be self-evident. This explains the pessimism of the famous final sentence of the first edition: "the Cosmos of Duty is thus really reduced to a Chaos: and the prolonged effort of the human intellect to frame a perfect ideal of rational conduct is seen to have been fore-doomed to inevitable failure." Later editions were less pessimistic. Having made the claim above Sidgwick swiftly claims that he is not to be taken as suggesting that "it would become reasonable for us to abandon morality altogether" (CC.5.2/508). But, it has to be pointed out, it is hard to see how Sidgwick can claim that it would be unreasonable to do so. As Sidgwick says, we might still be prompted to do our duty on the basis of self-interest and sympathy, but when there is a conflict between self-interest and duty, it would have to be decided (in practice) by the weight of the "non-rational impulses" in play.

In the final paragraph of the *Methods*, Sidgwick raises the question of whether the very fact that some hypothesis is required to avoid a contradiction in an important area of thought is itself reason for accepting that hypothesis. Once again, however, he refers the issue elsewhere, this time to "general philosophy." Despite the fact that the outcome of his ethical project depends fundamentally on this issue, there is no extended discussion of it in his other works. This is especially odd given the coherentist elements already in place in Sidgwick's moral epistemology. Sidgwick was leaving the development of his project to posterity, and the words of his friend F. W. H. Myers, written shortly after

Sidgwick's death, seem especially apposite:

> [H]e pointed to a definite spot; he vigorously drove in the spade; he upturned a shining handful; and he left us as his testament, *Dig here*.[11]

I am most grateful to the editors for detailed and extremely helpful comments on earlier versions of this chapter.

Notes

1. References in the text are to book, chapter, section, paragraph, and page of the 7th edition. So the reference in the second sentence of the first section below refers to book 3, chapter 2, section 3, paragraph 1, page 228.

2. See Sidgwick, *Essays*; and Sidgwick, *Outlines*.

3. Aristotle, *Nicomachean Ethics*, 1103b26–29.

4. Bradley, *Ethical Studies*, p. 81.

5. G. E. Moore, *Principia Ethica*, pp. 83–85.

6. Green and Grose, "Introduction," p. 7.

7. Parfit, *Reasons and Persons*, p. 388.

8. W. D. Ross, *The Right and the Good*, pp. 18–19.

9. Spencer, *The Data of Ethics*, chapters 15–16, especially pp. 268–275.

10. L. Stephen, *The Science of Ethics*, chapter 9, sections 12–15.

11. See Myers, *Fragments*, p. 108; cited in Schultz, *Henry Sidgwick*, p. 719.

5 Utilitarianism in the twentieth century

Krister Bykvist

Introduction

In the twentieth century, both the practice and the theory of utilitarianism were developed extensively. First, utilitarian thinking continued to have a pervasive influence on law, as well as political and economic policy. For example, one obvious reason for building a strong welfare state was that it actively promotes the well-being of the citizens. But utilitarian ideas were also applied to more specific moral questions, such as abortion, euthanasia, suicide, charity aid, and animal farming and experimentation. For example, Peter Singer's 1975 book on animal ethics, *Animal Liberation*, which highlighted the callous way we treat animals in medical research and factory farming, had a huge impact on private and public debates about animal ethics. What Singer took to heart was the utilitarian concern, first articulated by Jeremy Bentham in the eighteenth century, that "the question is not, Can they *reason*? nor, Can they *talk*? but, Can they *suffer*?"[1] As evidence of this impact, it is enough to point out that downplaying the importance of animal suffering is now often seen as a form of speciesism akin to racism and sexism.

Second, most of all what characterized utilitarian theorizing in the twentieth century was the aim for *greater precision*. A lot of research was devoted to making utilitarianism precise enough to dispel possible misunderstanding and permit rigorous, often formal, arguments. Early accounts of utilitarianism often lacked this level of precision. For instance, Hutcheson's infamous formulation of utilitarianism (later picked up by Bentham) says that "that Action is best, which procures the greatest Happiness for the greatest Numbers," which suggests that we should give independent weight to the sheer number of recipients of happiness, no matter whether this affects the total amount of happiness.[2] Bentham wrote that by utilitarianism is meant "that principle which approves or disapproves of every action whatsoever, according to the tendency which it appears to have to augment or diminish the happiness of the party whose interest is in question," which seems to suggest that an action should be approved of more strongly if it tends to produce more happiness, and not necessarily that we ought to do what will produce most happiness.[3] Mill defined utilitarianism as the principle that says that "actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness," without giving us any clear guidance on how to understand the crucial phrases 'right in proportion' and 'tend to promote'.[4]

The greater precision of utilitarian theories in the twentieth century was enabled by an analysis of each theory into its more fundamental parts. So, for instance, act utilitarianism, according to which an action ought to be done just in case its outcome

contains a sum total of well-being that is greater than that which is contained in the outcome of any alternative action, was broken down into two parts: Act Consequentialism and Sum-Ranking Welfarism.

> *Act Consequentialism:* You ought to perform an action if and only if its outcome would be better than the outcome of any alternative action available to you.
> *Sum-Ranking Welfarism:* An outcome x is better than another y if and only if the total sum of individual welfare contained in x is greater than that in y.

Act Consequentialism needs further refinement, since it does not have much content unless the crucial notions of 'alternative action' and 'outcome' are elucidated. While the early utilitarians were often happy to just talk about promotion of overall happiness without making it especially clear what promotion amounts to, twentieth-century utilitarians clarified the notion of promotion by taking great pains at explicating the notions of 'alternative action' and 'outcome'.[5] There was also a thorough discussion of indirect versions of utilitarianism, which require an indirect promotion of overall well-being. One such indirect version is rule utilitarianism, which tells the agent to choose an action that falls under a rule that would have the best consequences, if generally followed (see Chapters 6 and 7 for more on act versus rule utilitarianism).

Sum-Ranking Welfarism also raises a number of issues that were addressed at some length in the twentieth century. First of all, there was a lot of discussion of the correct substantive account of well-being (for more on substantive accounts of well-being, see Chapters 10 and 11). Whereas traditional utilitarianism identified well-being with pleasure, a major theme of utilitarian theorizing in the twentieth century was a move beyond the narrow confines of this kind of hedonism. In the hands of philosophers, decision theorists, and economists, hedonism was replaced by other, more inclusive, theories of well-being.

One such theory identifies well-being with preference satisfaction. When combined with Act Consequentialism and Sum-Ranking Welfarism, the result is *preference act utilitarianism*, which asks us to promote the satisfaction of our preferences.[6] On this view, what matters is not just that people feel good, but that they get what they want. Versions of this kind of utilitarianism quickly found a home in economics departments, where it has long been taken for granted that consumers are sovereign in the sense that each person is the best judge of his own good. Initially, it was assumed that the relevant preferences and judgments were revealed by people's choices in the market (or could even be identified with these choices), but, more recently, "happiness" economists, such as Richard Layard, have come to think that the relevant attitudes are revealed by people's own assessments of how satisfied they are with their lives.[7] On this approach, the importance of simple preference satisfaction is therefore played down in favor of "life-satisfaction."

Another, more remote, descendant of traditional utilitarianism is the theory that tells us to promote *objective* well-being. What matters, according to this theory, is not just that we feel good, or have our preferences satisfied, but that we have such things as good friends, freedom of choice, and an ability to perfect our nature by artistic, intellectual, and athletic achievements.[8]

Even if we set aside substantive questions about well-being, there is a further question of how to aggregate individual well-being. Sum-Ranking Welfarism tells us that we should aggregate by simply adding up individual levels of well-being. This principle is part of the core of twentieth-century utilitarianism. Despite extensive disagreements about the correct accounts of well-being, alternative actions, outcomes, and value promotion, all utilitarians endorsed Sum-Ranking Welfarism as the correct aggregation procedure when assessing the overall value of outcomes. Furthermore, it is this principle that was often seen as especially problematic by the critics of utilitarianism, since it entails that it can be better to sacrifice one person for the sake of small benefits for many others. Since Sum-Ranking Welfarism is such a crucial utilitarian principle and many of the other theoretical developments of twentieth-century utilitarianism are dealt with elsewhere in this volume, I shall devote the main part of this chapter to this principle. I shall first break down Sum-Ranking Welfarism into a set of axioms or principles, which together provide an exact characterization, and then briefly explain what is at stake in accepting each of them. In doing this, I will be applying the axiomatic approach to utilitarianism that was part of the core research in welfare economics and social choice theory in the twentieth century and remains so. Next, I shall present John Harsanyi's equiprobability argument for utilitarianism, which has its roots in the axiomatic approach to utilitarianism. Finally, I will present R. M. Hare's role-reversal argument for utilitarianism, which shares with Harsanyi's argument the crucial idea that moral decisions should be based on the sympathetic identification with others.

Sum-Ranking Welfarism

Many different formal characterizations of Sum-Ranking Welfarism were developed during the twentieth century, but I shall only discuss one fairly recent characterization. To simplify, I will avoid the formal and mathematical details. I will also simplify the discussion by focusing exclusively on outcomes in which the same people exist. That is, I will not consider formal characterizations of Sum-Ranking Welfarism for variable-population contexts, where the number or the identity of people varies from one outcome to another. (See Chapter 16 for utilitarianism in variable-population contexts.)

The first principle to be assumed in the axiomatic characterization of Sum-Ranking Welfarism is Cardinal Interpersonal Comparability, according to which different people's well-being gains and losses can be compared. On this view, it makes sense to say, for instance, that my gain in well-being is greater than your loss in well-being. This is a

crucial assumption, since Sum-Ranking Welfarism is not meaningfully defined unless we can compare different people's gains and losses. We do not need to assume that the well-being *levels* of different people can be compared; for Sum-Ranking Welfarism to be meaningfully defined it is enough that gains and losses in well-being can be compared across different people. However, some objections to utilitarianism assume that we can compare levels of well-being across different people, so that we can say, for instance, that I am better off than you. In order to accommodate these objections, I will also assume Level Comparability in the following.

Two further innocent-looking principles that are assumed are:

Pareto Indifference: If outcomes x and y are equally good for everyone, then x is equally as good as y.
Independence: The comparative value of two outcomes x and y depends only on what is happening in x and y, and not on what is happening in other outcomes.

These two principles together entail Welfarism, which states that only welfare factors can make a difference to the value of outcomes.[9] Welfarism thus excludes information about desert, merit, freedom, justice, and rights (assuming that well-being is not even partly constituted by these factors). Since Welfarism follows from these two principles, you need to reject either Pareto Indifference or Independence (or both), if you want to reject Welfarism.

If we assume Welfarism, we can represent each outcome by a list (or vector) of well-being values u, which we can call a *well-being profile*:[10]

$$u = (u_1, u_2, \ldots, u_n),$$

where u_1 stands for the first individual's well-being, u_2 stands for the second individual's well-being, and so on. So, for instance, the outcome in which Jane's well-being is 10, John's 2, and Lisa's 20 could be represented by the well-being profile (10, 2, 20).

In this framework, Sum-Ranking Welfarism can be formulated more precisely thus:

Well-being profile $u = (u_1, u_2, \ldots, u_n)$ is at least as good as well-being profile $v = (v_1, v_2, \ldots, v_n)$ just in case $u_1 + u_2 + \ldots + u_n$ is at least as great as $v_1 + v_2 + \ldots + v_n$.

Not all welfarist principles are sum-ranking, obviously. For instance, a theory of general maleficence that claims that we make things better by making people *worse off* is welfarist in the broad sense employed here. General maleficence is excluded if we add the constraint that we always make things better by making everyone better off. Or, to put it in terms of well-being profiles:

Weak Pareto: If each value in well-being profile u is greater than the corresponding value in well-being profile v (i.e., u_1 is greater than v_1, u_2 is greater than v_2, and so on), then u is better than v.

Weak Pareto is not enough to guarantee Sum-Ranking Welfarism, however. Some welfarist principles that satisfy Weak Pareto give different weights to the well-being of different people. For example, one welfarist principle would give everyone's well-being some positive weight but give more weight to the well-being of one particular individual. But utilitarians are famous for giving every person's well-being equal weight, or as Bentham is often supposed to have said, "everybody to count for one, nobody for more than one."[11] This kind of impartiality is captured by the following principle.

Anonymity: If well-being profile u is a permutation of the values in well-being profile v, then u is as good as v.

Anonymity says that a rearrangement of the values in a well-being profile always gives you an equally good well-being profile. Anonymity thus implies that the following well-being profiles are equally good: (2, 4, 6), (2, 6, 4), (4, 2, 6), (4, 6, 2), (6, 2, 4), (6, 4, 2).

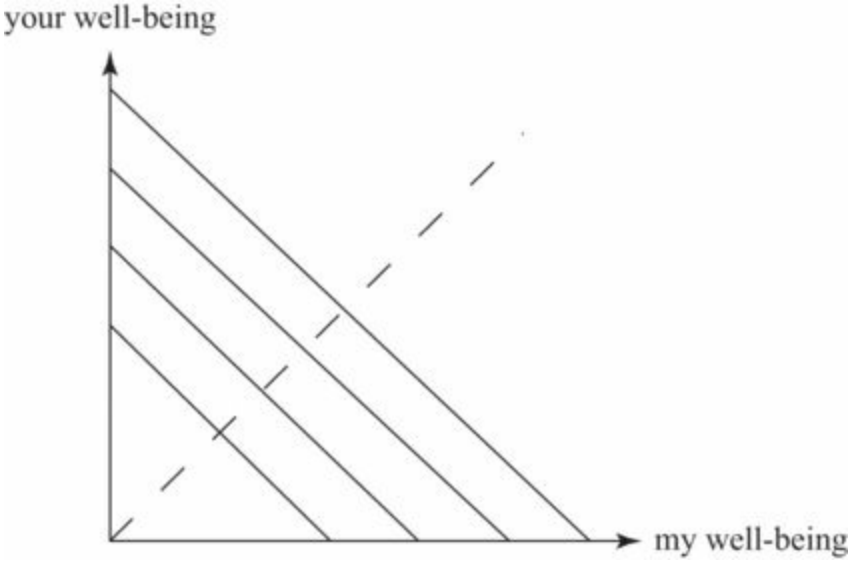
We are not home yet, because not all anonymous welfarist principles that satisfy Weak Pareto are sum-ranking. Consider, for example, Maximin, according to which a well-being profile u is at least as good as another profile v just in case the well-being of the worst-off in u is at least as good as the well-being of the worst-off in v. Maximin would agree that all the profiles listed in the previous paragraph are equally good, since the level of well-being of the worst-off does not change from one profile to another (though the *identity* of the worst-off does change). So, it is clear that Maximin is an anonymous welfare principle. Maximin also satisfies Weak Pareto, since if you make *everyone* better off you also make sure that the well-being of the worst-off is greater.

It can be shown that we get all the way to Sum-Ranking Welfarism if we replace Anonymity with a stronger principle that says that any gain for one person is exactly balanced in value by an equally sized loss for another person. Or, more formally:

Transitional Equity: Well-being profile u is as good as the well-being profile we get if we increase some individual well-being value in u by amount k and decrease any other individual well-being value in u by the same amount k (where k is any real number).[12]

For example, this principle implies that (2, 6, 4) is as good as (2, 6 + k, 4 – k), for any k. So, in particular, (2, 6, 4) is as good as (2, 8, 2), where k = 2. More generally, Transitional Equity implies that any transfer of a certain amount of well-being from one person to another does not affect the overall value.

It can be shown that, given Transitional Equity and Weak Pareto, two well-being profiles with the same sum-total of well-being are equally good, and a well-being profile with greater total well-being is always better.[13] Consider the following graph, which illustrates this result for the two-person case.



Each point represents a well-being profile (u_1, u_2) where u_1 is my level of well-being and u_2 is yours. Points on the dashed right-leaning diagonal represent well-being profiles in which you and I have exactly the same well-being. Points on any given left-leaning diagonal have the same sum total of well-being as each other. Transitional Equity guarantees that all points on any given left-leaning diagonal are equally good. Weak Pareto guarantees that the points on the dashed right-leaning diagonal are not equally good: the more to the right you move along the diagonal, the better profile you get, since for each move we both get better off. Together these principles guarantee that a profile is better than another just in case it has a greater sum total of well-being.

This axiomatic way of presenting Sum-Ranking Welfarism has many virtues. First of all, the axiomatic approach makes it perfectly clear what is at stake when deciding to accept or deny it. If you want to deny Sum-Ranking Welfarism, you cannot at the same time accept Cardinal Interpersonal Comparability, Pareto Indifference, Independence, Weak Pareto, Anonymity, and Transitional Equity. At least one of these principles has to be rejected.

Another, related, virtue of the axiomatic approach is that we can now more clearly decide which aspects of Sum-Ranking Welfarism we like and which we do not like. Some would deny Cardinal Interpersonal Comparability and thus deny that it makes sense to compare gains and losses across different people. But it should be noted that, no matter which stand we take on utilitarianism, we often assume that it makes sense to make such comparisons. For instance, we assume that the loss in well-being I experience

when my dentist pulls out all my teeth without an anesthetic is greater than the gain in well-being you experience when you eat a small candy.

Others would ask us to give up Welfarism (that is, either Pareto Indifference, Independence, or both) and allow non-welfare factors to make a difference to the value of outcomes. It should be noted here that one has to be careful not to reject Welfarism on the basis of an overly narrow conception of well-being. As pointed out in the introduction, a utilitarian need not be wedded to a simple-minded version of hedonism, according to which only the quantity of pleasure matters for well-being.

Given Welfarism, Weak Pareto seems pretty uncontroversial since it says that things get better if *everyone* is made better off. But one could object that it does not take into account inequality in terms of *gaps* in well-being between different individuals. Even if we make sure that everyone benefits, some might benefit much more than others. This "mind the gap" version of egalitarianism has its own problems, however, the most famous one being the leveling-down objection. If the gap between different people's well-being matters, it seems difficult to avoid the conclusion that you can make the world better (at least in one respect) by making everyone worse off as long as you create a more equal distribution of well-being. To see this, consider this schematic example:

	profile 1	profile 2
my well-being	19	–10
your well-being	1	–10

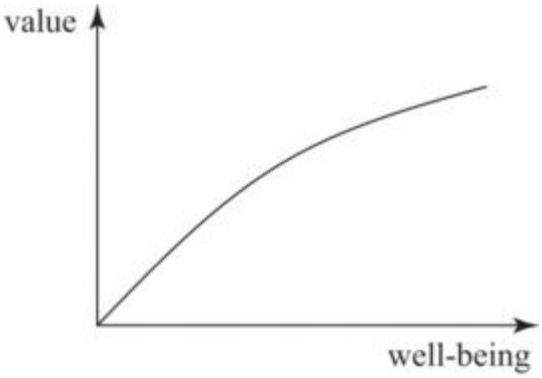
If you are an egalitarian and think welfare gaps always matter, you have to say that the second profile is better than the first – at least in one respect, "no inequality" – even though everyone is worse off in the second profile (indeed, we can assume that we are both suffering in the second profile).

Anonymity is also pretty uncontroversial, at least if we already assume Welfarism. Indeed, it is often seen as one defining feature of moral impartiality, and many non-utilitarian welfarist theories accept this condition, including the Maximin principle, presented above, that gives all weight to the worst-off person.

The final principle, Transitional Equity, is definitely one of the most controversial principles. It may seem very innocent, for who could object to giving equal weight to different people's equally sized gains and losses? But appearances are misleading. If we assume that we can compare well-being levels across people, Transitional Equity loses much of its attractiveness. For example, it seems very plausible to say that a transfer of a certain amount of well-being from a better-off person to someone who is worse off counts as an improvement so long as the better-off person is still better off and the

worse-off is still worse off, but this runs counter to Transitional Equity. (This kind of transfer is often called a Pigou–Dalton transfer in the social choice literature.)

In reaction to this, many have proposed a *prioritarian* alternative to utilitarianism, according to which priority is given to the worse-off.[14] Unlike Maximin, which gives all weight to the worst-off, prioritarianism gives each person's well-being some positive weight, but more weight is given to the well-being of worse-off people. More exactly, the idea is that the weight depends on the person's *absolute* level of well-being, so the weight assigned to your well-being level does not depend on the well-being levels of other people. However, the lower a person's absolute level of well-being, the more weight is assigned to that person's level of well-being. We can illustrate this with the following graph:



The horizontal axis represents a person's absolute level of well-being and the vertical axis represents the value that is assigned to that level of well-being. The upward slope of the graph shows that all benefits count. The downward bend shows that less weight is given to a benefit if it is received when one has a higher level of well-being. A simple version of prioritarianism could then be stated thus:

> *Prioritarianism:* Well-being profile u is at least as good as well-being profile v if and only if the sum of all the weighted well-being values in u is at least as great as the corresponding sum in v.

More formally, $u = (u_1, u_2, \ldots, u_n)$ is at least as good as $v = (v_1, v_2, \ldots, v_n)$ if and only if $w(u_1) + w(u_2) + \ldots + w(u_n)$ is at least as great as $w(v_1) + w(v_2) + \ldots + w(v_n)$, where $w(\bullet)$ is the function plotted in the figure above (a strictly increasing concave function).

Prioritarianism will not only welcome transfers of well-being from the better-off to the worse-off and thus deny Transitional Equity; it will also avoid the leveling-down objection, since prioritarianism implies that things get worse if some people are made worse off and no one is made better off.

It is important not to be misled by the catchy slogan "Give priority to the worse-off"

here. It may seem that by giving more weight to worse-off people, prioritarianism is biased in favor of some people over others because they are worse off than others, and that does not look like impartiality. Compare: if I give more weight to the well-being of the rich and famous, I seem to be showing bias toward some people over others because of the way they compare to others.

It is a mistake, however, to think that prioritarianism is a partial view that favors certain people because of how they compare with others. Note first of all that since prioritarianism satisfies Anonymity, it is concerned only with the well-being levels of individuals and not with the identity of the persons who enjoy these levels of well-being. In a very clear sense, then, prioritarianism honors impartiality. Furthermore, how much value a life has does not depend on how it fares *in comparison* to other lives; it only depends on the absolute well-being level of the life. I can be better off than you in one situation, and worse off than you in another, but if my absolute level stays the same in both situations, its weight will also be the same.

A genuine problem for prioritarianism is that it is not clear how the weights should be determined. Exactly how much weight should be given to a person at a certain absolute well-being level? If the different weights given to the worse-off and the better-off differ only marginally, then the resulting theory will come pretty close to utilitarianism. If they differ radically, then too little weight is given to the better-off.[15]

Harsanyi's equiprobability argument

Harsanyi gives two arguments for Sum-Ranking Welfarism: the aggregation argument and the equiprobability argument.[16] Both have been discussed extensively since the first versions of these theorems were published in the 1950s. I will only discuss the equiprobability argument here, since it is much easier to see how it is supposed to support Sum-Ranking Welfarism (in fact, many doubt that the aggregation argument supports a full-blooded version of Sum-Ranking Welfarism).[17] It also has some interesting connections with Hare's argument, which I will discuss later.

Harsanyi's equiprobability argument for Sum-Ranking Welfarism draws a connection between the aggregation of well-being across people and rational preference in risky choice situations. To explain this argument (a much simplified version of the argument, I should add), consider the following toy example. Suppose we have two outcomes, o_1 and o_2, where o_1 is associated with the welfare profile (1, 6, 6) and o_2 with (6, 3, 3). Harsanyi now asks you to compare the relative merits of o_1 and o_2 from a moral point of view, which he characterizes as the point of view you would take if you were *impartial*, *sympathetic*, and *rational*.

To make sure that your preference is *impartial* and cannot be influenced by any selfish or personal considerations, Harsanyi thinks you need to be ignorant (or imagine

being ignorant) about which person you are in o_1 and in o_2. More exactly, he assumes that you need to think, for each outcome, that there is an *equal chance* that you end up being in any person's place in that outcome. So, for example, for each of outcomes o_1 and o_2, you need to think that there is a $\frac{1}{3}$ chance of being in the first individual's place, a $\frac{1}{3}$ chance of being in the second's place, and a $\frac{1}{3}$ chance of being in the third's place.

To make sure that your preference is *sympathetic*, you need to imagine not just being in the other person's objective situation – having her career, income, and health, for example – but also being in her subjective situation, including having her tastes, values, and preferences. So, you need to imagine not just being in the other person's shoes, but being in her shoes with her feet! Note that you are only asked to imagine being in the other person's objective and subjective circumstances, which can be shared by different persons. You are not asked to imagine the impossibility of being *identical* to someone else.

Harsanyi maintains that by this act of sympathetic imagination your preference for being in a certain person's place in a certain outcome – what Harsanyi calls an *extended* preference – will coincide with the well-being of the person in that outcome. More exactly, he maintains that the strength of your extended preference for being in that person's place will coincide with the amount of well-being of that person in that place.

The comparison of outcomes o_1 and o_2 is thus conceptualized as a comparison of the following two *lotteries*:

lottery 1:			
probability	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
your preference strength	1	6	6
lottery 2:			
probability	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
your preference strength	6	3	3

In the last move, Harsanyi invokes the rationality part of the moral point of view, and he assumes that rationality requires you to prefer one lottery to another just in case the first has a greater expected utility than the other (where utility is assumed to represent

preference strength). The expected utility of a lottery is a weighted average of the utilities of the possible outcomes, where the weights are the probabilities assigned to the outcomes. The expected utility calculations of lotteries 1 and 2 are thus:

$$\text{your expected utility of lottery 1} = \frac{1}{3}(1) + \frac{1}{3}(6) + \frac{1}{3}(6) = \frac{13}{3}$$

$$\text{your expected utility of lottery 2} = \frac{1}{3}(6) + \frac{1}{3}(3) + \frac{1}{3}(3) = \frac{12}{3}$$

Hence, your rational preference is for lottery 1 over lottery 2. Since this is the rational preference you would have if you were impartial and sympathetic, we can conclude that your moral preference must be for outcome o_1 = (1, 6, 6) over o_2 = (6, 3, 3). This result can be generalized:

> Outcome x with well-being profile $(x_1, x_2, \ldots, x_n)$ is *morally* preferred to outcome y with well-being profile $(y_1, y_2, \ldots, y_n)$ just in case $\frac{1}{n}(x_1)+\frac{1}{n}(x_2)+\ldots+\frac{1}{n}(x_n)$ is greater than $\frac{1}{n}(y_1)+\frac{1}{n}(y_2)+\ldots+\frac{1}{n}(y_n)$, that is, since $\frac{1}{n}$ cancels out, just in case $x_1 + x_2 + \ldots + x_n$ is greater than $y_1 + y_2 + \ldots + y_n$.

So, one outcome is morally preferred to another if the first contains a greater sum total of well-being than the other. If we add that an outcome is better if it is morally preferred, we have Sum-Ranking Welfarism.

This is an intriguing argument and it has triggered a lot of critical discussion. (I will put aside the more technical and less accessible parts of this discussion.)

One objection is that Harsanyi does not seem to have shown that we can compare different people's well-being gains and losses. This objection is important because, as pointed out earlier, utilitarianism is not well-defined unless such comparisons can be made. One reply is to concede that this is not shown but simply assumed in the argument. This need not be so questionable, however, since, as mentioned earlier, you do not need to be a utilitarian to find Cardinal Interpersonal Comparability plausible. Another much more controversial reply is to say that comparisons of well-being gains and losses between people are simply *constituted* by extended preferences. Suppose that I have an extended preference for being in person A's place in outcome x rather than being in her place in outcome y, and also an extended preference for being in person B's place in outcome y rather than being in her place in outcome x. If my extended preference concerning A is stronger than my extended preference concerning B, then this *means* that A gains more than B loses when we move from y to x.

A more fundamental criticism against the argument is presented by Rawls, who claims

that Harsanyi misrepresents impartiality.[18] He agrees with Harsanyi that in order to model impartiality the agent should put herself behind a "veil of ignorance" by imagining not knowing which position she will occupy in society, but Rawls's veil is thicker than Harsanyi's. Rawls thinks that in order to model impartiality properly we should ask you to assess the outcomes from the perspective of *complete uncertainty* of which person you are in the outcomes. From this perspective, you know the well-being values of the people in the outcomes (or people's primary well-being resources, which Rawls wants to focus on), but you are ignorant of the probability of being in any person's place in an outcome. So, in the example above, you have to choose between two uncertain prospects, (1, 6, 6) and (6, 3, 3), and you know only the following: if you go for the first prospect, there are three possibilities for you – a life at well-being level 1, one at 6, or a different one at 6 – and, if you go for the second, three different possibilities – a life at 6, one at 3, or a different one at 3. Rawls then argues that under complete uncertainty it is rational to prefer to "play it safe" and thus judge prospects by their worst possibility. So, in the example above, it is rational to prefer the prospect (6, 3, 3) to the prospect (1, 6, 6), since the worst possibility of the first prospect is a life at well-being level 3 and the worst possibility of the second is a life at level 1. The final step is to move from rational preference over prospects to betterness of outcomes, which in the case at hand means that o_2 = (6, 3, 3) is judged as better than o_1 = (1, 6, 6). The result is not Sum-Ranking Welfarism but Maximin, which tells you to maximize the well-being of the worst-off.

Who is right, Harsanyi or Rawls? This depends on how impartiality is best modeled, but this is not the place to discuss this thorny issue. It is worth pointing out, however, that the rationality principle Rawls endorses is highly problematic, *if* we can use probabilities, for it would then tell us to maximize the value of the worst possibility even if the worst possibility is extremely unlikely. However, the disagreement between Rawls and Harsanyi is exactly about when probabilities can and should be used. Whereas Rawls insists that we should only use them when they are grounded in some empirical evidence about the chances of the possibilities, Harsanyi is happy to use *subjective* probabilities – degrees of confidence – even in cases where there is no such empirical evidence accessible to the agent. More specifically, in these cases, Harsanyi thinks that it is rational to assign the same subjective probability to all possibilities.

Another radical objection to Harsanyi's argument is to question whether what is morally preferred in Harsanyi's stipulated sense is also better. To echo a question raised by Brian Barry, if I am initially disinclined to believe that it would be better if I sacrificed a lot for the sake of the benefits of others – perhaps I am the person who stands to lose if (1, 6, 6) is realized rather than (6, 3, 3) – why should I change my mind if I am told that I would prefer the sacrifice if I were in a very different situation in which I only knew that I had an equal chance of being in any person's place?[19]

Since a thorough answer to this question would lead us into a discussion of the nature of morality, I will only give the beginnings of a reply. First of all, it is important to note

that this objection would also apply to Rawls's theory since he agrees with Harsanyi that what we would prefer in a situation that is very different from the real one is also what is morally better (or *just*, as Rawls would say). Furthermore, it seems pretty clear that the last step in the argument could be defended if one adopted a *contractualist* account of morality, according to which moral facts about what we ought to do and what is better than what are constituted by facts about what we would all agree on if we were in an ideal situation of deliberation. Different contractualist theories differ on how to understand the notion of an ideal situation. Some claim that we only need to be rational and well informed. Others would add more substantive conditions. Harsanyi can be interpreted as providing a more substantive conception of what it means to be in an ideal situation. To be in an ideal situation is to be not just rational, but also impartial and sympathetic. On this contractualist construal of Harsanyi's view, the fact that one outcome x is better than another y is constituted by the fact that we would all prefer x to y in an ideal situation in which we are impartial, sympathetic, and rational.

This contractualist construal highlights another problem with the argument, however. Harsanyi assumes that we would all *agree* in our preferences if we were impartial, sympathetic, and rational. But this is doubtful. Suppose that I am a violinist and you a city banker, and that we are asked to compare my life as a violinist to your life as a city banker. What guarantees that we both would end up having the same extended preference concerning these lives after we have fully imagined living either life? Since I am a committed violinist, I may very well prefer my life as a violinist to your life as a city banker, while you, a committed city banker, may have the opposite preference, or so it seems.[20]

Hare's role-reversal argument

Unlike Harsanyi, Hare starts from explicitly metaethical assumptions.[21] First of all, Hare assumes that moral judgments express *prescriptions*. When I sincerely utter "You ought to apologize" I am not describing the way the world is, I am prescribing an action, namely your act of apologizing, and to prescribe an action is to prefer that the action be performed. Furthermore, moral prescriptions are *universalizable* in the sense that when I prescribe an action for one situation, I am committed to prescribing the same action for all situations that share the same universal features. For example, when I prescribe that you apologize to me in a situation in which you have offended me, I am committed to prescribing that I apologize to you in a hypothetical situation in which the roles have been reversed and I have offended you. Finally, these prescriptions are *overriding* in the sense that they trump other kinds of motivational factors. For example, my moral commitment to apologize when I have offended you trumps my self-interested motivation to keep quiet and not draw attention to my faults.

In addition to these metaethical assumptions – "Prescriptivity," "Universalizability,"

and "Overridingness," as they are often labeled – Hare assumes a conceptual principle about knowledge about one's preferences in hypothetical situations. This principle, which he does not give a name but I will call 'The Principle of Hypothetical Reflection', says that knowledge about one's preference *in* a hypothetical situation entails a matching preference *about* that hypothetical situation.

> *The Principle of Hypothetical Reflection:* If I know that I would prefer A, with a certain strength, in a hypothetical situation S, then I now prefer, with the same strength, that A is done in S.[22]

For example, if I know that I would strongly prefer not to be offended in a hypothetical situation in which you are offending me, I *now* prefer that I not be offended in that hypothetical situation.

Hare's final assumption is that *practical rationality* requires you to resolve conflicts between your own preferences by balancing your preferences by their strength.[23] If you have conflicting preferences regarding the outcomes of two actions A and B, some preferences for A's outcome and some preferences for B's, then it is rational for you to prefer action A to B just in case the sum of intensities of your preferences for A's outcome over B's is greater than the sum of intensities of your preferences for B's outcome over A's.

Hare aims to show that an agent who satisfies these assumptions will end up having preferences that agree with the ranking provided by utilitarianism, more specifically, *preference utilitarianism*, according to which one action is better than another just in case the first action brings about a greater sum of preference satisfaction.

To see how the argument works, let us go through an example that Hare himself uses.[24] Suppose that you want to park your car in a spot that is occupied by someone else's bicycle. Suppose further that you have a strong preference for the removal of the bicycle, and the cyclist has a weak preference for not removing the bicycle. To fix our ideas, let us assume that the situation looks like this (unlike Hare, I have added numbers to represent the preference strengths).

actual situation, first stage:

people	preference	strength
you	for the removal of the bicycle	6
the cyclist	against the removal of the bicycle	2

Universalizability entails that whatever you prescribe for this situation, you also need to prescribe for the following hypothetical situation in which the roles are reversed.

hypothetical situation:

people	preference	strength
the cyclist	for the removal of the bicycle	6
you	against the removal of the bicycle	2

Since you know that you would prefer the bicycle not to be removed if you were in the cyclist's exact circumstances, the Principle of Hypothetical Reflection entails that you *now* prefer that, in the hypothetical situation in which you are in the cyclist's exact circumstances, the bicycle is not removed. Furthermore, this principle implies that the strength of your actual preference for this hypothetical situation will be the same as the strength of your hypothetical preference in that hypothetical situation. So, the actual situation is changed into the following.

actual situation, second stage:

people	preference	strength
you	for the removal of the bicycle	6
you	against the removal of the bicycle in the hypothetical situation in which the roles are reversed	2

In the third stage, we apply practical rationality, and you balance your preference for removal against your preference against removal. Since the preference for removal is stronger, you end up preferring overall the removal of the bicycle. In the final step, you then universalize this preference and also acquire a preference for the removal of the bicycle in the hypothetical situation in which the roles are reversed. It is this preference for the removal of the bicycle, in both actual and hypothetical cases, that is expressed by your moral judgment that you ought, all things considered, to remove the bicycle. We seem to be home, because your final preference exactly agrees with the preference-utilitarian ranking of the actions.

This is how Hare's argument seems to proceed at a first glance. But there is a crucial gap in this argument: there is no conflict between your preferences in the second stage.[25] One of your preferences is for the removal of the bicycle, but the other preference is against the removal of the bicycle in the *hypothetical* case in which the roles are reversed. Since the preferences concern different situations, there is no conflict. Compare: There is no conflict between my current preference for eating fermented herring and my current preference for not eating fermented herring in a hypothetical situation in which my taste is different and I find it disgusting.

This gap has to be filled. One suggestion is that we *tentatively* universalize the preference for the hypothetical case and see whether it survives a conflict with the preference for the actual case.[26] So, we start by tentatively universalizing the preference against removing the bicycle in the hypothetical case and end up with a preference of strength 2 against removing the bicycle in the actual case. Since this preference is weaker than the preference for removing the bicycle, we end up with a preference for the removal of the bicycle.

The problem with this account is that it works only for bilateral cases. It does not work for cases where two or more weak preferences *together* outweigh a strong preference. Suppose, for instance, that we have four cyclists and that each has a preference of strength 2 that the bicycle not be moved. Suppose you take over each of these preferences by going through the process of hypothetical reflection four times, so you end up with four actual preferences for four different hypothetical situations, one situation for each cyclist. Since each of your preferences for the hypothetical situations would, on its own, lose the competition with your stronger preference for the actual removal of the bicycle, you would end up having a preference for the removal of the bicycle – a preference that does *not* agree with the preference-utilitarian ranking.

Another proposed solution, one that avoids this problem, is that we follow Harsanyi in adopting an equiprobability model and imagine that we have an equal probability of being the motorist and the cyclist.[27] We imagine being in each person's place and taking over their preferences and, finally, we apply expected utility theory to our newly acquired preferences. The result is that we prefer to remove the bicycle, since removing the bicycle will maximize expected utility, for $\frac{1}{2}(6)+\frac{1}{2}(0)$ is greater than $\frac{1}{2}(0)+\frac{1}{2}(2)$.

This would fill the gap but at a significant cost for Hare: the universalizability principle is no longer applicable. Remember that the principle tells you to consider what *would* be the case, if the roles were reversed. You are not asked to pretend that you have an equal chance of being either the motorist or the cyclist. You are asked to assess a counterfactual situation in which you and the cyclist have swapped roles.[28]

So, it is still unclear how to best fill the gap in Hare's argument.[29] But no matter how this issue is resolved, we can ask whether the premises of the argument survive scrutiny.

Prescriptivity is perhaps the most contentious assumption, since it amounts to non-cognitivism in metaethics, the view that moral judgments do not purport to describe any moral facts.

Universalizability seems less controversial. In fact, you can deny Hare's non-cognitivism and think that coherent moral *beliefs* must be universalizable: if I believe that I ought to do A to you, then I must also believe that you ought to do A to me, were our roles reversed. On this cognitivist account of morality, moral beliefs must be universalizable because it is a general *truth* that if an action ought to be done in a situation, then it ought to be done in any situation that shares the same universal features.

The Principle of Hypothetical Reflection is far from uncontroversial. It seems pretty clear that it cannot be a *conceptual* truth. Whether my actual preferences for a hypothetical situation coincide with my hypothetical preferences in that situation depends on whether I care about my hypothetical self. But it is not a conceptual truth that I care for my hypothetical self.

Of course, one could reply that I *should* care about my hypothetical self. But the principle seems to fail even if it is seen as a normative principle. Suppose you know that your attitudes would be corrupted in the hypothetical scenario. Perhaps you know that they would be deeply irrational or monstrously immoral. Why should we demand that your actual preference coincide with the irrational or immoral preferences you would have in the hypothetical situation?

In reply, one could try to qualify the principle so that it only applies to cases of uncorrupted preferences. The case about the motorist and the cyclist is exactly such a case. So, Hare's argument may work at least for these cases.

Concluding remarks

The high level of theoretical sophistication in utilitarian thinking has made it possible, for the first time, to exactly identify the utilitarian commitments. While this is a great theoretical achievement in its own right, it may seem to come at a significant cost. It seems to turn ethics into a formal discipline that loses touch with the true nature of our moral experiences. To characterize utilitarianism in terms of precise axioms or principles may seem to be a purely technical exercise with very little relevance to our moral life and traditions. It is therefore important to stress that both Harsanyi's and Hare's arguments are not just abstract theoretical constructions. Remember that they both identify moral decisions with decisions based on sympathetic identification. That morality has to do with sympathetic identification is one of the most popular moral ideas in the history of humanity, and it is endorsed by all major world religions, including Christianity, Islam, Judaism, and Hinduism. One particularly catchy version of this idea is the Golden Rule, which asks you to do to others what you want them to do to you. Harsanyi and Hare can be seen as being in the business of refining the golden rule so that it tells you to do to

others what you *would* want them do to you, if the roles were reversed and you were in their exact objective and subjective circumstances. Conceived in this way, utilitarianism is not a number-crunching, cold-hearted theory; it is a theory that tries to articulate one crucial aspect of our common moral experiences: the moral importance of putting yourself in another person's place and seeing things from her point of view. It is not surprising, then, that utilitarianism is still one of the main contenders in normative ethics.

Notes

1. Bentham, *IPML*, p. 283, n. b.

2. Hutcheson, *An Inquiry into the Original*, p. 125.

3. Bentham, *IPML*, p. 12.

4. J. S. Mill, *Utilitarianism*, *Collected Works*, vol. X, p. 210.

5. See, for instance, Bergström, *The Alternatives and Consequences of Actions*; and Carlson, *Consequentialism Reconsidered*.

6. See, for instance, Hare, *Moral Thinking*; Harsanyi, "Morality and the Theory of Rational Behaviour"; and Singer, *Practical Ethics*.

7. Layard, *Happiness*.

8. See, for instance, Griffin, *Well-Being*, pp. 64–72.

9. More exactly, they entail Welfarism if we are also prepared to assess outcomes with any possible combination of welfare and non-welfare information. For a derivation of Welfarism, see Blackorby, Bossert, and Donaldson, *Population Issues*, pp. 59–60, or d'Aspremont and Gevers, "Equity and the Informational Basis of Collective Choice," pp. 199–209.

10. Blackorby, Bossert, and Donaldson, *Population Issues*, p. 62. See also

d'Aspremont and Gevers, "Equity and the informational basis of collective choice," pp. 199–209.

11. J. S. Mill, *Utilitarianism*, *Collected Works*, vol. X, p. 257.

12. This principle is similar to the principle "Incremental Equity" which is discussed in Blackorby, Bossert, and Donaldson, *Population Issues*, pp. 118–119.

13. For a similar proof, see Blackorby, Bossert, and Donaldson, *Population Issues*, pp. 118–119.

14. The remainder of this section is based on Bykvist, *Utilitarianism*, pp. 70–72.

15. For more on prioritarianism, see Holtug, "Prioritarianism."

16. Harsanyi, "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking"; and Harsanyi, "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility."

17. For a good summary and assessment of this debate, see Weymark, "A Reconsideration of the Harsanyi–Sen Debate on Utilitarianism."

18. Rawls, *A Theory of Justice*, pp. 183–192.

19. Barry, *Theories of Justice*, pp. 334–335.

20. Broome, *Weighing Goods*, p. 55.

21. Hare, *Moral Thinking*, pp. 21–22 and pp. 55–57.

22. Hare, *Moral Thinking*, pp. 95–96.

23. Hare, *Moral Thinking*, pp. 109–110.

24. Hare, *Moral Thinking*, p. 109.

25. This gap was spotted by, among others, Schueler, "Some Reasoning about Preferences" and Persson, "Universalizability and the Summing of Desires."

26. Suggested by Rabinowicz and Strömberg, "What If I Were in His Shoes?"

27. Suggested by Persson, "Universalizability and the Summing of Desires."

28. Pointed out by Rabinowicz and Strömberg, "What If I Were in His Shoes?"

29. For some interesting recent proposals, see Rabinowicz and Strömberg, "What If I Were in His Shoes?"; and Rabinowicz, "Preference Utilitarianism by Way of Preference Change."

6 Act utilitarianism

Ben Eggleston

The basic idea

The definition of act utilitarianism

In nearly every part of the world, there is moral opprobrium attached to the idea of a doctor ending a patient's life, even if the patient sincerely requests it because he has a terminal, debilitating, and painful disease. Efforts to relieve such patients' pain are widely regarded as humane, but active euthanasia is widely condemned both ethically and legally. Suppose that, despite these prohibitions, a doctor gives a lethal injection to such a patient. Depending on the circumstances, the doctor might be subject to general excoriation, the revocation of her license, and even criminal prosecution. But has she done anything wrong?

Act utilitarianism, like other forms of utilitarianism, approaches questions of this kind by holding that morality is ultimately a matter of overall well-being. What distinguishes act utilitarianism from other, rival, forms of utilitarianism is the extremely direct and straightforward way in which it specifies this basic utilitarian idea. It holds, quite simply, the following:

> *Act utilitarianism:* An act is right if and only if it results in at least as much overall well-being as any act the agent could have performed.

In other words, in any situation, an agent acts rightly if she maximizes overall well-being, and wrongly if she does not. In the example given above, if the lethal injection promoted overall well-being at least as much as any act the doctor could have performed, then it was right, according to act utilitarianism. And if it did not, it was wrong.

Act utilitarianism, indeed utilitarianism more generally, is both broad and narrow in ways that are sometimes surprising to people when they first encounter the view. It is remarkably broad because of its account of whose well-being matters to the moral value of an act. A natural thought about well-being and morality is that only the well-being of people directly affected by an act can influence the moral value of that act. But act utilitarianism holds that all well-being – experienced by any being (human or otherwise), in any place (near or far), at any time (whether in the present or the remote future) – matters to the moral value of any individual act. For example, to the extent that Plato's dialogues continue to bring pleasure (or displeasure) to twenty-first-century students of philosophy, the precise moral value of acts committed more than two millennia ago is still

evolving. Thus, in the definition of act utilitarianism, in the phrase 'overall well-being', the breadth of the adjective cannot be overstated.

Though act utilitarianism is remarkably broad in the way just mentioned, there is another way, also having to do with well-being, in which act utilitarianism is strikingly narrow. Few would question the thought that the moral value of an act depends at least in part on whether it makes people better or worse off: only a truly bizarre moral theory would hold that well-being does not matter, morally. But act utilitarianism goes to the opposite extreme, holding that *only* well-being matters, morally. Whatever other properties a particular act might have – e.g., that it was a felony, or was an instance of disloyalty, or was done from selfish motives – these properties do not have any independent relevance to the moral value of the act. They might indicate ways in which the act has affected or will affect overall well-being, but they do not matter in and of themselves. This narrowness of act utilitarianism is arguably the most distinctive aspect of the view and, indeed, the utilitarianism tradition generally: that it focuses on this one item that is widely regarded as relevant to the moral values of acts – their effects on overall well-being – and declares that nothing *else* is relevant to the moral values of acts.

Three further clarifications

An artificially schematic but nonetheless useful way to think about act utilitarianism is in terms of the choice situation that a given agent faces at a given time. Suppose that, in a particular choice situation, an agent can choose to perform any of, say, seven acts: A_1, A_2, A_3, . . ., A_7. Now suppose that, for any act A_i, we define $W(A_i)$ as the world that would result if act A_i were performed, so that in the case at hand, we have $W(A_1)$, $W(A_2)$, $W(A_3)$, . . ., $W(A_7)$. We then rank these worlds according to the amounts of overall well-being they respectively contain. In the resulting ranking, either there will be just one of these worlds in first place or there will be two or more of these worlds tied for first place. If just one of these worlds is in first place, then the corresponding act is not only right but obligatory, with every other act being wrong. If several of these worlds are tied for first place, then no particular act is obligatory, but it is nonetheless obligatory for the agent to choose only from among the acts corresponding to the tied-for-first-place worlds, and any such act is right. As in the simpler case, every other act – every non-utility-maximizing act – is wrong.

This way of thinking about act utilitarianism is particularly useful in dispelling the objection that act utilitarianism is impractical or incoherent because its goal of utility maximization can never be achieved. This objection is based on the claim that regardless of the act any agent performs at any time, there will always be more work for him and others to do in the future toward the maintenance and production of well-being. Although this claim is obviously true, it does not present a problem for act utilitarianism, since act utilitarianism (like all prominent forms of utilitarianism) is compatible with the idea that promoting well-being is a never-ending enterprise rather than a discrete task that some

agent might have the opportunity to bring to completion. An agent's duty at any given time, according to act utilitarianism, is not to act so that the resulting world has as much overall well-being as a world can have, but just to act so that the resulting world has as much overall well-being as any world that could have resulted from the acts that were among the agent's options at the time of acting. In other words, the idea of maximization that act utilitarianism involves is the idea of maximizing over the agent's set of options, not the idea of maximizing in the sense of leaving no increases to be achieved subsequently.

A second idea meriting further clarification is that of a world's overall well-being. When we think about a world, in all its spatial vastness and the entire duration of its existence, what does its overall well-being consist of? For utilitarianism, it is just the sum of the well-being had by the entities in that world that are capable of having well-being. (In most utilitarian theories, these are the organisms that are capable of feeling pleasure and pain.) Each such entity will have some total, lifetime well-being – positive, it is to be hoped, but possibly negative – and the sum of those, for all of the world's creatures, is that world's overall well-being.

Conceiving of overall well-being in this way helps to pre-empt a misunderstanding than can arise from the phrase 'the greatest happiness for the greatest number', which dates from the eighteenth century[1] and remains, in common parlance, synonymous with utilitarianism. On its face, this phrase suggests that an act should not only produce as much happiness as possible, but should also produce happiness for as many people as possible. That makes this phrase problematic as a criterion of right action, since it is often the case that the most beneficial act is different from the act that will spread the benefit most widely (since, in many choice situations, a small set of people has much more at stake than the rest of humanity does). In contrast, when overall well-being is conceived simply as the sum of individuals' well-being (as explained above), the 'for the greatest number' part of the phrase proves otiose.[2] Maximizing overall well-being might often result from the act that benefits the most people, but even in that case the act is right (according to act utilitarianism) simply because it maximizes overall well-being, not because it benefits the most people. We may conclude, with Russell Hardin, that "No philosopher should ever take the dictum of the greatest good for the greatest number seriously except as a subject in the history of thought."[3]

Finally, we saw above that according to act utilitarianism, nothing other than overall well-being matters to the moral value of an act. For example, the fact that an act is a crime, or results from a vicious character trait, does not make it wrong; moreover, such a fact does not detract from its moral value at all, according to act utilitarianism. By the same token, act utilitarianism entails that the moral value of an act does not depend, at all, on whether the act complies with any kind of moral rule (other than the act-utilitarian rule of "Maximize well-being"). This is important because the concept of a rule is often regarded as integral to the concept of morality. For example, morality is often understood

as the rules for the regulation of behavior that are generally accepted (in the agent's society, typically),[4] or the rules that are generally accepted that satisfy some ethical criteria,[5] or the rules that ought to be generally accepted, regardless of whether they are currently accepted.[6] Potential examples of rules meeting one or more of these criteria are the prohibition against active euthanasia, mentioned above, and the requirement that the owners of pets keep them reasonably comfortable.

According to act utilitarianism, the fact that an act would comply with, or would violate, a rule that meets any criterion such as those just mentioned is irrelevant to its moral value: all that matters is how the act would affect overall well-being, relative to how alternative acts would affect overall well-being. This is not to say, of course, that in practice act utilitarianism is blind to the existence and potential usefulness of moral rules. The existence of moral rules can affect the way an act benefits or harms people; for example, in a society with a moral rule against the cremation of the bodies of revered elders, such an act would have different consequences than in a society that accepts cremation as a valid practice. Moreover, it is consistent with act utilitarianism to hold that, as a matter of psychological and sociological fact, the existence of certain moral rules, in a given society or throughout the world, can be useful for the promotion of overall well-being, because they are an effective device for the restraint and coordination of behavior. The catch is that such rules would not, according to act utilitarianism, have any actual bearing on the moral value of acts done in that society, or anywhere. Act utilitarianism's simultaneous repudiation of moral rules (as irrelevant to the moral value of acts) and embrace of them (as potentially beneficial tools) is arguably the most subtle and complex aspect of act utilitarianism, and we will return to this topic in the last section of this chapter.

Historical and contemporary context

Act utilitarianism has a long history, having been espoused in landmark utilitarian treatises such as Jeremy Bentham's *An Introduction to the Principles of Morals and Legislation* (1789),[7] Henry Sidgwick's *The Methods of Ethics* (1st edn., 1874),[8] and G. E. Moore's *Principia Ethica* (1903).[9] For some or all of these authors it may have been, to some extent, merely the default form of utilitarianism rather than a conscious choice, since it was not explicitly and influentially formulated as a particular kind of utilitarianism until the 1950s. That decade, however, saw the emergence of rule utilitarianism as a well-defined alternative to act utilitarianism – presented as important both historically, for the interpretation of Mill's *Utilitarianism* (1861), and as a substantively plausible view[10] – resulting in a correspondingly heightened precision in the delineation of act utilitarianism as one specific option within the broader utilitarian school of thought. The label 'act-utilitarianism' seems to have entered the philosophical literature in 1959,[11] and within two years there appeared the unhyphenated variant, which has become the more common

term.[12]

To further contextualize act utilitarianism within the utilitarian school of thought, let us define rule utilitarianism more precisely:

> *Rule utilitarianism:* An act is right if and only if it would be permitted by a system of rules whose general acceptance would result in at least as much overall well-being as would the general acceptance of any system of rules.

Rule utilitarianism affirms act utilitarianism's claim that rightness is conceptually dependent on overall well-being, but denies act utilitarianism's claim that the dependence is direct, or immediate: instead, it holds that the dependence is indirect, because it is mediated by rules. Now, it is possible to affirm rule utilitarianism's claim that the dependence is mediated, but deny rule utilitarianism's claim that rules are what do the mediating. One might, for example, privilege motives instead:

> *Motive utilitarianism:* An act is right if and only if it would result from the motives whose general possession would result in at least as much overall well-being as would the general possession of any motives.[13]

Along these lines one can envision conscience utilitarianism,[14] virtue utilitarianism,[15] and so on. Such views are often labeled "indirect" forms of utilitarianism (though this term is problematic because it is also used to describe a kind of act utilitarianism, as explained below). Rule utilitarianism is the most thoroughly developed and discussed of these, and is the main rival of act utilitarianism within contemporary utilitarian thought. I will discuss rule utilitarianism further below, and Dale Miller discusses it more thoroughly (and sympathetically) in his chapter in this volume (Chapter 7). Meanwhile, it is worth noting that act utilitarianism's claim that rightness is *directly* conceptually dependent on well-being is one of its most important characteristics.

Some theorists who accept the directness of act utilitarianism object to a different component of it: that of *maximization*, in its requirement that acts maximize overall well-being. Most such theorists recommend, instead, the concept of "satisficing," first presented by Michael Slote as holding that "an act might qualify as morally right through having good consequences, even though better consequences could have been produced in the circumstances."[16] (Slote took the term 'satisfice' from the writings of the economist Herbert Simon.[17] Although it is often assumed to be a portmanteau of 'satisfy' and 'suffice', Slote notes that "it is a Scotticism for 'satisfy'.")[18] Proponents of satisficing forms of utilitarianism generally claim that its demands are more reasonable than are those of maximizing forms of utilitarianism; for example, an act that results in a great deal of overall well-being but does not happen to maximize well-being might be right according to a satisficing form of utilitarianism but would, obviously, be wrong

according to act utilitarianism.[19]

In response to the satisficing proposal, defenders of maximizing sometimes express skepticism that the view can really amount to anything other than a nuanced form of maximizing. For example, Robert Goodin writes that "maximization under constraints of time and information costs" is "the best sense I can make of" the view.[20] And when satisficing is construed as a genuine alternative to maximizing, defenders of the latter insist that it cannot be permissible to intentionally choose consequences that one acknowledges to be all-things-considered worse than consequences that one might choose instead. In short, many follow Philip Pettit in claiming "that [a defender of satisficing] is committed to unmotivated sub-maximization and that this is profoundly irrational."[21] The relative merits of maximizing and satisficing continue to be debated.[22]

The final component of the contemporary context of act utilitarianism concerns the many questions to which there is no definite act-utilitarian answer, so that act utilitarians can (and do) differ among themselves about how best to answer them. Perhaps the most prominent of these questions is about the nature of well-being, which is discussed in the chapters by Chris Heathwood and Ben Bradley (Chapters 10 and 11). Act utilitarianism requires the maximization of well-being, but is compatible with various conceptions of well-being. Similarly, act utilitarianism is compatible with various answers to the question of whether the moral value of an act depends on its actual effects on overall well-being or the effects that could reasonably have been expected when the act was performed. (We often say to people who unwittingly do harm, "You couldn't have known.") These possibilities, and others, are discussed in the chapter by Elinor Mason (Chapter 9). As a final example, act utilitarianism is compatible with various answers to the question of whether what is to be maximized is the total quantity of well-being or the average of the levels of well-being had by the entities that are capable of having well-being. The significance of this issue, and the most important arguments that bear on it, are explained in the chapter by Tim Mulgan (Chapter 16). Above, act utilitarianism is presented in terms that explicitly or implicitly refer to an act's *actual* effects on *total* well-being, but this is just for expository convenience: a sufficiently flexible formulation would be cumbersomely complex.

Supporting arguments

The range of arguments that can be given in support of act utilitarianism is remarkably diverse, and many of the arguments are complex. But most of them are elaborations of one of two basic strategies that can be presented here in a stylized form.[23]

Respecting individuals' interests

The first strategy for justifying act utilitarianism starts with individuals' interests, and

regards morality as primarily concerned with resolving conflicts between those interests. Some such conflicts have obvious resolutions. For example, some individual might have an interest in owning a bank account that someone else owns instead. But some such conflicts are not so one-sided. For example, some commuters might have an interest in the widening of a particular road while nearby residents prefer the status quo. This strategy for justifying act utilitarianism sees an individual's interests as constituting his or her well-being. So, conflicts between individuals' interests are seen as conflicts between individuals' well-being. Thus, morality is primarily concerned with what should be done when increasing one individual's well-being entails decreasing (or just declining to increase) another individual's well-being.

To resolve such conflicts, this strategy for justifying act utilitarianism holds that the strength of individuals' claims to the maintenance and improvement of their well-being is proportional to the *magnitudes* of the *changes* in their well-being that are under consideration in a particular case.[24] Thus, in a two-person conflict, if the first person stands to experience a large increase in well-being and the second person's well-being will be reduced only slightly, the first person's claim would have more moral weight. This strategy also holds that, in cases affecting more than two people, the strengths of multiple individuals' claims should be combined. Consider, for example, a case resembling the previous one, except that instead of just one person facing a slight reduction in well-being, there are several. If they are numerous enough, their claims would have more moral weight than the first person's claim, even though none of their claims, taken individually, would have more moral weight than the first person's claim.

Now, it so happens that the foregoing way of resolving conflicts between individuals' well-being is equivalent to the maximization of the total quantity of well-being: if every conflict is resolved in accordance with the claims corresponding to the largest amounts of well-being at stake, no outcome containing less well-being than another possible outcome will ever be chosen. Thus, on this view, the moral imperative of respecting individuals' interests is made more determinate, and summed up, by the act-utilitarian principle of maximizing overall well-being.

Sum-ranking welfarist act consequentialism

The second strategy for justifying act utilitarianism begins with a focus on states of affairs, rather than people (or other individuals), and it holds that some states of affairs are better than others. It then makes this idea more determinate by embracing two additional theses. One of these holds that only well-being contributes to the goodness of a state of affairs:

> *Welfarism:* The value of a state of affairs is positively related to, and determined by nothing other than, the well-being it contains.[25]

Now, this thesis is compatible with several mutually exclusive theses about how the value of a state of affairs is determined by the valuable things (such as well-being) it contains. One of these is concerned with equality, holding that the value of a state of affairs is positively related to, and determined by nothing other than, how equally the valuable things (whatever they may be) are distributed in that state of affairs.[26] Another is concerned with minimizing disadvantage, holding that the value of a state of affairs is positively related to, and determined by nothing other than, the quantity of valuable things that is enjoyed by the individual who has the smallest quantity of it.[27] A third is concerned with maximizing the total quantity of what is valuable. This, of course, is the thesis that contributes to the present strategy for justifying act utilitarianism:

> *Sum-ranking:* The value of a state of affairs is positively related to, and determined by nothing other than, the total quantity of value it contains.[28]

Combining the theses of welfarism and sum-ranking yields the view that the value of a state of affairs is positively related to, and determined by nothing other than, the total quantity of well-being it contains. This view is, obviously, very close to act utilitarianism.

But this view is only about states of affairs, not acts. So the justificatory strategy under consideration embraces one further thesis, about the way the value of every act depends on its consequences:

> *Act consequentialism:* An act is right if and only if its consequences are at least as good as the consequences of any act the agent could have performed.

This thesis, combined with welfarism and sum-ranking, completes this justification for act utilitarianism. This justification is, then, a *sum-ranking welfarist act-consequentialist* one.

Contrast and historical examples

Although the two strategies obviously have much in common, they are fundamentally different. The first strategy takes individuals' interests as fundamental and takes morality to be primarily concerned with resolving conflicts between those interests. The concept of maximization comes in fairly late in the proceedings, almost as a mathematical accident. If this strategy's principle for how to assess the relative strengths of competing claims were altered, the resulting view could fail to be a maximizing one.

In contrast, the second strategy is an essentially maximizing one: it takes morality to be primarily concerned with maximally promoting valuable states of affairs. Correspondingly, individuals, and their interests, come in somewhat later – if not accidentally, then certainly more subordinately. If this strategy's principle for what

determines the value of a state of affairs were altered, morality might not have anything to do with individuals, and their interests, at all.

Perhaps because of the second strategy's impersonal, "top down" orientation, the first strategy has been preferred by more of the major figures in the history of utilitarianism. Variations of it are arguably deployed by Bentham ("every individual in the country tells for one; no individual for more than one")[29] and Mill ("To do as one would be done by, and to love one's neighbour as oneself, constitute the ideal perfection of utilitarian morality"),[30] as well as by the twentieth-century and contemporary theorists John Harsanyi ("a social welfare function ought to be based . . . on the utility functions (subjective preferences) of *all* individuals, representing a kind of 'fair compromise' among them"),[31] R. M. Hare ("We are led to give weight to the preferences of all the affected parties . . . in proportion to their strengths"),[32] and Peter Singer ("when we make ethical judgments . . . we weigh interests").[33]

Although the first strategy has been the more popular one, the second strategy has had its influential exponents as well. For example, Sidgwick is famous for suggesting that the moral point of view is "the point of view . . . of the Universe," and he reports that "it is evident to me that as a rational being I am bound to aim at good generally . . . not merely at a particular part of it."[34] Sidgwick's cosmological aspirations are shared by Moore, who focuses on "the greatest possible amount of good in the Universe" and who writes that "the primary and peculiar business of Ethics" is not, for example, the resolving of people's conflicts, but "the determination [of] what things have intrinsic value and in what degrees."[35] Moreover, the second strategy may enjoy greater prominence than its historical frequency would suggest because of its neat factoring of utilitarianism into distinct components[36] and because of the rise, in recent decades, of consequentialism as a focal point within the discipline of moral philosophy.[37]

Objections

The basic idea of act utilitarianism has a certain obvious appeal: well-being is a fine thing, and of course folks should have more of it rather than less. There are, however, several important objections to act utilitarianism. These objections have been prominent throughout the history of the view as well as presenting ongoing challenges for contemporary theorists.

Impractical to implement

Perhaps the most straightforward objection to act utilitarianism is that it is impractical to implement in everyday decision-making, due to the difficulty of predicting the effects on well-being of all of the possible acts that make up a given choice situation. Mill

anticipates this concern, at least in part, and replies that during "the whole past duration of the human species" people "have been learning by experience the tendencies of actions."[38] For example, experience has shown that people tend to become corrupted by power.[39]

But tendencies are not enough: act utilitarianism holds that the moral value of an act depends on all of its effects on well-being, however atypical, far-flung, or delayed they may be. And because many of an act's effects on well-being are unforeseeable, act utilitarianism seems to require, for its competent implementation, an impossible degree of foresight. For example, people aiding strangers risk inadvertently putting those strangers in harm's way.[40] And Hitler's ancestors could not have known, when they engaged in procreation, that their actions would eventually be among the causes of the Holocaust.[41] Such examples suggest that act utilitarianism may be impractical to implement, compromising its initial appeal.

Harmful if implemented

A second objection also focuses on act utilitarianism's implementation, but moves beyond the issue of bare feasibility to claim that such implementation would have severe consequences for ordinary decision-making, social order, and virtue. When a high-placed government official of eighteenth-century England called the utilitarian principle a "dangerous" one, Bentham playfully pretended to fail to understand how it could ever be "not consonant to utility to consult utility" – before explaining, with equal zest, that what his contemporary feared, quite rightly, were the reforms utilitarianism would prescribe for institutions that bestowed and perpetuated unmerited privilege.[42]

But the implementation of act utilitarianism, critics claim, would have consequences more troubling than those that worried eighteenth-century elitists. First, if people were to set aside the common-sense morality that prevails today and were to adopt the practice of making decisions according to the act-utilitarian standard of maximizing overall well-being, decision-making itself would become cripplingly tedious and time-consuming.[43] Second, there would be much more selfish behavior, since predictions of consequences are often fraught with uncertainties and people have a well-known tendency to resolve such uncertainties in ways that agree with their own interests.[44] Third, coordination would break down, since people would expect one another not to stick to previously made plans, but to regard every choice point as a fresh opportunity for maximizing.[45] Fourth, there would also be breakdowns of socially beneficial virtues such as honesty, the keeping of promises, and the special ties constitutive of love and friendship, since such virtues require people to act on principles that are not fully captured by the act-utilitarian goal of maximizing overall well-being.[46] In sum, the result would be little more than slow and selfish decision-making conducted by uncoordinated moral deficients.

Such a prospect would, of course, reflect badly on any moral theory. But it is especially discrediting to act utilitarianism, critics argue, since that theory's core ideal is the maximization of overall well-being and the prospect just sketched is, among its many failings, an utter debacle on that score. On these grounds, act utilitarianism is often said to prohibit its own implementation, and to be "self-defeating."[47]

Immoral implications

The most serious and influential objection to act utilitarianism concerns the moral judgments that act utilitarianism entails for particular cases of moral decision-making. According to this objection, there are countless cases – reflecting diverse aspects of morality – in which act utilitarianism entails judgments that are questionable or utterly unacceptable. Such cases, it is claimed, show that act utilitarianism misconstrues, or just runs roughshod over, many important aspects of morality.

One such aspect of morality includes the various special obligations that people are often thought to have. Promises, as well as figuring in the preceding section, make for an apt example here: people are often thought to have special obligations in virtue of promises they have made, but act utilitarianism is said to fail to give promises their proper moral weight. This claim is developed in a classic discussion from W. D. Ross:

> Suppose . . . that the fulfillment of a promise to *A* would produce 1,000 units of good for him, but that by doing some other act I could produce 1,001 units of good for *B*, to whom I have made no promise . . . We should, I fancy, hold that only a much greater disparity of value between the total consequences would justify us in failing to discharge our *prima facie* duty to *A*. After all, a promise is a promise, and is not to be treated so lightly as the theory we are examining would imply.[48]

The special obligations that stem from promises are not the only ones that act utilitarianism is said to neglect. Others include the special obligations that people have to other people in virtue of what those other people have earned, or deserve; and the special obligations that people have to their family and friends. All of these cases, it is said, show that morality is not just a matter of maximizing overall well-being.

A second aspect of morality that act utilitarianism is said to violate has to do with treating individuals justly: act utilitarianism is said to be too ready to impose grave harms on some people in order to provide benefits to others. Consider this case provided by T. M. Scanlon:

> Suppose Jones has suffered an accident in the transmitter room of a television station. Electrical equipment has fallen on his arm, and we cannot rescue him without turning off the transmitter for fifteen minutes. A World Cup match is in progress, watched by many people, and it will not be over for an hour. Jones's

injury will not get any worse if we wait, but his hand has been mashed and he is receiving extremely painful electrical shocks.[49]

Scanlon asserts that we should rescue Jones immediately, regardless of how many people are watching the match. But act utilitarianism implies that if the viewers are numerous enough, we should wait. In a similar vein, other theorists have argued that act utilitarianism condones the judicial punishment of innocent people, depending on the facts of the situation. Cases can be imagined in which the harm experienced by the innocent person is outweighed by other benefits, such as quieting social unrest or deterring other people from performing harmful acts.[50] Act utilitarianism's readiness to regard harms to some people as outweighed by benefits to others is part of the basis of John Rawls's famous claim that act utilitarianism "does not take seriously the distinction between persons."[51]

A third aspect of morality where act utilitarianism is said to go astray involves the sacrifices that it requires individuals to make in order to provide benefits to other people. Cases illustrating this issue are structurally similar to those concerned with treating individuals justly, except that the person experiencing the harm is the agent himself or herself, rather than another person. Paradigm cases concern the extent of the obligations of affluent people to donate money to poverty-relief programs. As it happens, such cases are asserted by proponents of act utilitarianism, as well as by opponents of the theory. The former argue that because of the moral imperative of promoting overall well-being, affluent people are obligated to donate much larger sums of money than is generally thought to be obligatory;[52] the latter argue that because act utilitarianism is so demanding, it must be wrong.[53] (The former's *modus ponens* is the latter's *modus tollens*.) Despite some proponents' candor about the demandingness of act utilitarianism, this aspect of the theory remains one of the main grounds on which critics claim that it has a distorted view of morality.

Indirect utilitarianism

Overview

Mindful of the foregoing objections, contemporary proponents of act utilitarianism tend to advance a particular form of the view that is often called "indirect utilitarianism" (though, as noted earlier, this term is also often used to refer to rivals of act utilitarianism such as rule utilitarianism). Although proponents of this view intend for it to overcome or mitigate all of the foregoing objections, the second objection provides an especially convenient point of entry into this view. In response to this objection – that the implementation of act utilitarianism would be harmful, making act utilitarianism self-defeating – defenders of act utilitarianism point out that this objection presupposes the

use of act utilitarianism as what might be called a *decision procedure*. That is, it presupposes that people use act utilitarianism as a procedure for deciding what to do in ordinary choice situations. And defenders of act utilitarianism then concede that act utilitarianism is not well suited to be used in that way, for precisely the reasons stated in the second objection. But they claim that act utilitarianism is a defensible moral theory nonetheless, because it offers the correct *criterion of rightness* – the correct account of what makes actions right and wrong. On this view, the fact that act utilitarianism is not well suited for use as a decision procedure reflects, at most, something unfortunate about the psychological and social costs of pursuing the aims of morality too directly, and not any failure on the part of act utilitarianism to provide a sound account of what ultimately determines the moral values of acts.

This, then, leaves act utilitarians with the question of what decision procedure to recommend. The principle underlying their answer is simple: for any given person, the ideal decision procedure is the one whose possession and employment by that person would maximize overall well-being. For most people, the ideal decision procedure is probably some variant of common-sense morality: a decision procedure giving considerable weight to values such as honesty, the keeping of promises, the special ties constitutive of love and friendship, and so on. Of course, the exact contours of the ideal decision procedure for any given person is a complicated empirical question involving all of the myriad considerations mentioned in the articulation of the objection about the harmfulness of implementing act utilitarianism. Whatever the exact contours of the ideal decision procedure turn out to be, indirect utilitarianism is characterized by (1) affirming act utilitarianism as the correct criterion of rightness and (2) regarding the ideal decision procedure to be the one that best advances the goal of maximizing overall well-being.[54]

It is important to avoid the misperception that the ideal decision procedure proposed by indirect utilitarianism is essentially act utilitarianism augmented with a collection of rules and guidelines carefully tailored to enable an agent who is consciously searching for the act that will maximize overall well-being to identify it, in the way that a shopper intent on buying the most delicious tomato might apply guidelines that recommended choosing tomatoes with a particular color, weight, firmness, or aroma. Rather, the ideal decision procedure being sketched here is one in which the agent values the goods mentioned above – honesty, the keeping of promises, the special ties constitutive of love and friendship, and so on – for their own sakes, even though act utilitarianism entails that these things matter merely as means to the promotion of overall well-being. Psychologically, such valuing might take the form of explicitly regarding certain rules as morally binding – perhaps the rules endorsed by rule utilitarianism – or might take the form of an unarticulated (but nonetheless firm) motive or disposition to act in certain ways in certain situations. In any case, whereas an agent using act utilitarianism as her decision procedure would unhesitatingly set any of the aforementioned goods aside when convinced that doing so would lead to the maximization of overall well-being, an agent with the ideal decision procedure would feel pangs of guilt at the prospect of setting any

of them aside – even when she is convinced that doing so would maximize overall well-being. A person's decision procedure is, in effect, her conscience, with all of the moral emotions that concept suggests. So, the ideal decision procedure proposed by indirect utilitarianism is not just a well-informed act utilitarianism. It is, rather, act utilitarianism complemented by other moral rules, motives, and dispositions. Although the resulting decision procedure contains elements with non-act-utilitarian content, it is recommended by act utilitarianism because of its favorable impact on overall well-being.

So, indirect utilitarianism rests on divorcing the notion of a criterion of rightness from the notion of a decision procedure, and maintaining that the correct criterion of rightness will not necessarily be advisable, or even self-endorsing, as a decision procedure. As a result, indirect utilitarianism contrasts interestingly with rule utilitarianism – the view that an act is right if and only if it would be allowed by (what is here called) the ideal decision procedure. Like rule utilitarians, indirect utilitarians regard act utilitarianism as self-defeating, in the sense described above. But whereas rule utilitarians ensure agreement between the correct criterion of rightness and the ideal decision procedure by, in effect, regarding the ideal decision procedure as constituting the correct criterion of rightness, indirect utilitarianism divorces the two notions in order to maintain act utilitarianism as the correct criterion of rightness.

Assessment

The merits of indirect utilitarianism are a subject of ongoing debate. In support of the view, one might attempt to rebut the objections surveyed earlier by claiming that indirect utilitarianism improves on direct act utilitarianism by being easier to implement, by being more beneficial when implemented, and by endorsing the having of moral commitments that closely match what are often regarded as important aspects of morality.

But indirect utilitarianism is vulnerable to various criticisms as well. First, one might dismiss, as irrelevant, what moral commitments a moral theory endorses *the having of*; what matters, one might claim, are the theory's *implications*, and on this score indirect utilitarianism can offer no improvement over direct act utilitarianism, since indirect utilitarianism's criterion of rightness is simply the principle of act utilitarianism.[55] Second, indirect utilitarianism seems to confirm rather than answer a longstanding additional objection to act utilitarianism – the objection that it is ineligible to serve as society's publicly affirmed morality.[56] This follows from the substantial overlap between the idea of society's publicly affirmed morality and the idea of a moral theory as a decision procedure. Given that indirect utilitarianism involves disavowing act utilitarianism as a decision procedure, it seems to thereby concede that it cannot well serve as society's publicly affirmed morality – or, at least, cannot well serve as the entirety of society's publicly affirmed morality.

Indirect utilitarianism is clearly more complicated than direct act utilitarianism, and it challenges several conventions of moral theory as traditionally practiced. It might,

however, be the most promising theoretical framework in which to embed the principle that morality is, fundamentally, simply a matter of maximizing overall well-being.

Notes

1. See Hutcheson, *An Inquiry into the Original*, p. 125 (though there the last word of the phrase is plural), and Bentham, *Comment and Fragment*, p. 393.

2. Bentham, *Deontology*, p. 309. This page is in Bentham's "Article on Government" in that volume.

3. Hardin, *Morality within the Limits of Reason*, p. 22.

4. Such a possibility is discussed (critically) in Gert, *The Nature of Morality*, p. 119.

5. See, for example, R. B. Miller, "Actual Rule Utilitarianism," p. 22; see also p. 7.

6. See, for example, Hooker, *Ideal Code, Real World*, p. 32; see also p. 144, n. 3.

7. Bentham, *An Introduction to the Principles of Morals and Legislation*, pp. 12–13 (in chapter 1).

8. Sidgwick, *The Methods of Ethics*, p. 411 (in bk. IV, ch. 1, § 1).

9. G. E. Moore, *Principia Ethica*, pp. 147–148 (in § 89). Moore's conception of what is to be maximized is not limited to well-being, so his view is not a form of act utilitarianism, strictly speaking. But it is close enough to have been influential in the development of utilitarian thought.

10. See the works mentioned by Smart, "Extreme and Restricted Utilitarianism," p. 344.

11. Brandt, *Ethical Theory*, p. 380.

12. Smart, *An Outline of a System of Utilitarian Ethics*, 1961 edn., p. 2.

13. This view is adapted from Adams, "Motive Utilitarianism," p. 470, where the phrase is suggested as a name for a particular view about the moral value of patterns of motivation.

14. See, e.g., Adams, "Motive Utilitarianism," p. 479; Brandt, *Facts, Values, and Morality*, p. 145; and Hooker, *Ideal Code, Real World*, p. 2, pp. 90–92, and pp. 131–132.

15. See, e.g., Crisp, "Utilitarianism and the Life of Virtue," p. 154.

16. Slote, "Satisficing Consequentialism," part I, p. 140.

17. Slote, "Satisficing Consequentialism," part I, pp. 141–142.

18. Slote, "Two Views of Satisficing," p. 27, n. 1.

19. Sinnott-Armstrong, "Consequentialism," section 6.

20. Goodin, *On Settling*, p. 34 and p. 83, n. 15.

21. Pettit, "Satisficing Consequentialism," part II, p. 172.

22. For recent discussions, see the papers collected in Byron, *Satisficing and Maximizing*; and B. Bradley, "Against Satisficing Consequentialism."

23. My division of arguments into these two kinds, along with some of the examples I cite, is indebted to Kymlicka, *Contemporary Political Philosophy*, pp. 32–37.

24. This is controversial, as explained in the discussion of the Transitional Equity principle in Krister Bykvist's chapter in this volume (Chapter 5).

25. Sen, "Utilitarianism and Welfarism," provides a thorough discussion of welfarism and an influential critique of it, particularly with a view to its use in utilitarianism. See

especially pp. 471–489.

26. Such a view is discussed at length in Dworkin, "What Is Equality? Part 1: Equality of Welfare."

27. This thought has obvious affinities with John Rawls's difference principle (Rawls, *A Theory of Justice*, pp. 75–83). This thought is developed differently in prioritarianism, a much-discussed cousin of utilitarianism for which the most-discussed source is Parfit, "Equality and Priority." Also see the earlier McKerlie, "Equality and Priority."

28. Sen, "Utilitarianism and Welfarism," pp. 468–471. Also see the discussion of sum-ranking welfarism in Krister Bykvist's chapter in this volume (Chapter 5).

29. Bentham, *Rationale of Judicial Evidence*, vol. VII, p. 334. Philip Schofield identifies this as the source of what Mill calls "Bentham's dictum": "everybody to count for one, nobody for more than one" (*Utility and Democracy*, p. 84, n. 25).

30. J. S. Mill, *Utilitarianism*, *Collected Works*, vol. X, p. 218.

31. Harsanyi, "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," p. 315. See also Harsanyi, "Morality and the Theory of Rational Behaviour," pp. 44–48.

32. Hare, "The Structure of Ethics and Morals," p. 187. See also Hare, *Sorting Out Ethics*, p. 145 (echoing Bentham's dictum).

33. Singer, *Practical Ethics*, p. 20.

34. Sidgwick, *The Methods of Ethics*, p. 382. The precise construction of the former phrase, in Sidgwick's text, is 'the point of view (if I may say so) of the Universe'.

35. Moore, *Principia Ethica*, p. 147 (in § 89) and p. 26 (in § 17).

36. See, for example, the excellent overview in Scarre, *Utilitarianism*, pp. 4–26.

37. Early visible examples of this trend include Scheffler, *The Rejection of*

Consequentialism (1982); and Railton, "Alienation, Consequentialism, and the Demands of Morality" (1984).

38. J. S. Mill, *Utilitarianism*, *Collected Works*, vol. X, p. 224.

39. J. S. Mill, *Considerations on Representative Government*, *Collected Works*, vol. XIX, p. 445.

40. McCloskey, "Utilitarianism: Two Difficulties," p. 62.

41. This example is adapted from pp. 344–345 of Lenman, "Consequentialism and Cluelessness," which offers a sophisticated presentation and discussion of this objection. Notable (and dissimilar) replies to Lenman include Dorsey, "Consequentialism, Metaphysical Realism and the Argument from Cluelessness"; and Burch-Brown, "Clues for Consequentialists."

42. Bentham, *Comment and Fragment*, p. 447; see also p. 516.

43. See, e.g., Railton, "Alienation, Consequentialism, and the Demands of Morality," pp. 153–154.

44. See Shaw, *Contemporary Ethics*, p. 146; and Hooker, *Ideal Code, Real World*, p. 143.

45. See Hodgson, *Consequences of Utilitarianism*, chapter 2; and Shaw, *Contemporary Ethics*, pp. 146–147.

46. See, e.g., Butler, *Works*, vol. II, pp. 190–192; Sidgwick, *The Methods of Ethics*, p. 136 and p. 405; Hodgson, *Consequences of Utilitarianism*, pp. 58–59 and p. 61; Stocker, "The Schizophrenia of Modern Ethical Theories," pp. 458–461; Wolf, "Moral Saints," pp. 427–430; and Kapur, "Why It Is Wrong to Be Always Guided by the Best," pp. 489–494.

47. See, e.g., Hodgson, *Consequences of Utilitarianism*, p. 3 and p. 60; and Parfit, *Reasons and Persons*, pp. 27–28 and pp. 40–41.

48. W. D. Ross, *The Right and the Good*, pp. 34–35.

49. Scanlon, *What We Owe to Each Other*, p. 235.

50. See, e.g., McCloskey, "A Non-Utilitarian Approach to Punishment," pp. 255–257; and Boonin, *The Problem of Punishment*, pp. 41–52.

51. Rawls, *A Theory of Justice*, p. 27.

52. See, e.g., Singer, "Famine, Affluence, and Morality"; and Unger, *Living High and Letting Die*.

53. B. Williams, "A Critique of Utilitarianism," pp. 108–118.

54. Some of the many notable works in the development of indirect utilitarianism are Bales, "Act-Utilitarianism"; Hare, *Moral Thinking* (especially chapter 2); Railton, "Alienation, Consequentialism, and the Demands of Morality"; and Crisp, "Utilitarianism and the Life of Virtue."

55. Elsewhere, I argue against this dismissal; see Eggleston, "Practical Equilibrium."

56. See, e.g., Sidgwick, *The Methods of Ethics*, pp. 489–490; and Hodgson, *Consequences of Utilitarianism*, p. 46. For contemporary responses to this objection, see Lazari-Radek and Singer, "Secrecy in Consequentialism"; and Eggleston, "Rejecting the Publicity Condition."

7 Rule utilitarianism

Dale E. Miller

What is rule utilitarianism?

Rule utilitarianism is the best-known and most frequently discussed alternative to act utilitarianism within the family of utilitarian moral theories. In *Forms and Limits of Utilitarianism*, David Lyons defines rule utilitarianism as the "theory according to which the rightness or wrongness of particular acts can (or must) be determined by reference to a set of rules having some utilitarian defense, justification, or derivation."[1] To elaborate on Lyons's definition, we might say that rule utilitarians hold (1) that actions' moral standings are determined by an "authoritative" moral code or set of moral rules (i.e., actions that are forbidden by the rules of this authoritative code are morally wrong, actions that are required by these rules are obligatory, and so on); and (2) that satisfying some utilitarian criterion is a necessary condition for rules to belong to the authoritative code.

It is instructive to contrast the roles of rules in rule and act utilitarianism. While a sophisticated version of act utilitarianism might direct agents to employ a "decision procedure" that incorporates a set of "summary rules" or "rules of thumb," it will not morally evaluate actions by reference to these rules. It will hold that an action that fails to maximize utility is wrong even if the agent selected it by a proper application of the appropriate decision procedure. For rule utilitarians, the authoritative moral code is not a mere decision procedure or heuristic device. Rather, it is the "moral standard" by which actions are to be morally evaluated.

The term 'rule utilitarianism' was coined in 1959 by one of the theory's most vigorous proponents, Richard Brandt.[2] J. J. C. Smart had referred to the view as "restricted utilitarianism" a few years earlier, but Brandt's is the name that stuck.[3] Rule utilitarianism is sometimes described as an example of "indirect utilitarianism,"[4] but this term is ambiguous. Sometimes, for example, 'indirect utilitarianism' is used specifically to describe sophisticated versions of act utilitarianism.[5]

Of course, the fact that rule utilitarianism was only given a name in the middle of the twentieth century does not mean that no one had conceived of it prior. On the evidence of his *Passive Obedience* we can classify George Berkeley as a "theological" rule utilitarian, as Colin Heydt notes in Chapter 1. Berkeley believes that God's commands are the ultimate source of our moral obligations. God, he argues, must will the well-being of mankind, and so must command us to promote this end. Yet we are not capable of knowing which of the myriad actions open to us at any given time would do so: "In

short, to calculate the events of each particular Action is impossible, and tho' it were not, would yet take up too much time to be of Use in the affairs of Life."[6] Thus we must have recourse to "the observation of certain universal, determinate Rules or Moral Precepts which, in their own Nature, have a necessary tendency to promote the Well-being of the Sum of Mankind."[7] Now were this all he said on the matter, one might suspect that Berkeley is a sophisticated act utilitarian who believes that God commands us to maximize utility on every occasion and who regards "determinate Rules or Moral Precepts" as heuristic devices to help us determine which specific actions would fulfill this command. Berkeley, though, goes on to make clear that this is not his view. Rather, "whatsoever practical Proposition doth to right Reason, evidently appear to have a necessary connexion with the universal Well-being included in it . . . is to be esteemed a Decree of God, and is consequently a Law to Man."[8] In other words, what God commands us to do just is to obey those rules compliance with which tends to promote well-being.

Berkeley may not be the only important historical figure to have held a rule-utilitarian view. Since 1953, with the publication of J. O. Urmson's "The Interpretation of the Moral Philosophy of J. S. Mill," many commentators have maintained that John Stuart Mill holds what amounts to a rule-utilitarian account of morality. While Henry West discusses Mill's moral philosophy in depth in Chapter 3, I will occasionally refer in passing to Mill below. No doubt the history of philosophy holds further thinkers who could be described as rule utilitarians or proto rule utilitarians, although I will not attempt to offer a comprehensive list. However, one line from a lecture of Kant's is worth mentioning because it sounds surprisingly like an endorsement of rule utilitarianism: "If we conduct ourselves in such a way that, if everyone else so conducted themselves, the greatest happiness would arise; then we have *so* conducted ourselves as to be worthy of happiness."[9] Of course, it would be a mistake to read too much into an isolated passage from a philosopher as notoriously difficult to interpret as Kant, and Jens Timmermann gives a systematic account of the relation between Kant and utilitarianism in Chapter 12.

The family of rule-utilitarian moral theories belongs to the even broader family of rule-consequentialist theories. Since the 1990s there has been a noticeable shift in the ethics literature away from discussions of rule utilitarianism specifically and toward discussions of rule consequentialism more generally. This shift may partly reflect a more general trend of giving closer consideration to non-utilitarian consequentialist theories, but it has much to do with the fact that the most prominent contemporary defender of rule consequentialism, Brad Hooker, has a non-utilitarian theory of the good that assigns intrinsic value not only to well-being but also to virtue and to equality in the distribution of well-being.[10] I will refer specifically to rule utilitarianism throughout most of this chapter, in keeping with the focus of this volume, but since Hooker's theory is identical to a rule-utilitarian view in every respect other than his theory of the good, I will still mention it at points to illustrate certain rule-utilitarian possibilities. I will also make

occasional mention of the work of another contemporary moral philosopher who has recently emerged as an advocate of a form of rule consequentialism, Derek Parfit.

John Rawls, in the much-cited "Two Concepts of Rules," distinguishes between act utilitarianism's summary rules and what might be called "practice rules"; the latter are rules that define "offices, roles, moves, penalties, defenses, and so on" within social practices and give these practices their "structure."[11] What's distinctive about practice rules is that they generate new forms of action that could not be performed in the rules' absence. In the absence of the rules that constitute the practice of promising, for instance, it would be impossible to make, or *a fortiori* to keep, a promise. Rawls's aim in drawing this distinction is to help rule utilitarianism answer objections that presuppose the summary conception of rules. However, his distinction is not exhaustive. A rule of conduct that requires us to give a certain percentage of our income to the less fortunate is not a practice rule, for instance, but if it is part of our moral standard then it is not a mere summary rule either. The rules that constitute a rule-utilitarian's authoritative moral code are not summary rules, but there is no reason to expect all or even most of them to be practice rules.

One final preliminary remark: The basic rule-utilitarian idea can be further developed in various ways, so rule utilitarianism itself constitutes a family of moral theories. In the following two sections I will discuss several dimensions on which rule-utilitarian theories can differ, but there are numerous others that I will not discuss in any depth. Rule utilitarians, like act utilitarians, can adopt different conceptions of well-being, frame their theories in terms of either actual or expected outcomes, and make either average or total utility the object of promotion. None of these choices raises any special issues for rule utilitarians in particular, though, so there is no need to dwell on them here.

Collective ideal-code rule utilitarianism

The most common approach to fleshing out the basic idea of rule utilitarianism has been to presuppose that the same moral code is authoritative for all of the members of some broad group and to say that this is the code whose "general adoption" among the members of that group would be utility maximizing, i.e., the code whose "general adoption utility" in that group is higher than that of any other code. This way of formulating rule utilitarianism yields what we can call "collective ideal-code rule utilitarianism": "collective" in virtue of the fact that it says that the same moral code is authoritative for all of the members of a broad group, and "ideal-code" in virtue of the fact that it says that the authoritative moral code is the one whose general adoption would be utility maximizing even if it is not in fact widely adopted within the group in question. Collective ideal-code rule utilitarianism is the most-discussed version of the theory today, and so it will be worth exploring at some length.

The notion of "general adoption" can itself be interpreted in several ways. We can

begin with the question of what it means to adopt a moral code. One possibility that suggests itself is to say that adopting a code just means complying with it perfectly, i.e., unfailingly obeying it. Some writers have argued that this interpretation of adoption apparently yields a rule-utilitarian theory that is "extensionally equivalent" with or collapses into act utilitarianism, agreeing with it about the moral standing of every action. The act-utilitarian Smart, for instance, writes that an "adequate rule-utilitarianism would not only be extensionally equivalent to the act-utilitarian principle . . . but would in fact consist of one rule only, the act-utilitarian one."[12] And the rule-utilitarian Brandt claims that even if the possibility of this one-rule code is disallowed on the grounds that it is "unfair to the rule-utilitarian view," there will still be some "enormously long and enormously complex" moral code whose prescriptions would be "identical with the acts prescribed by the act-utilitarian principle." The code would presumably be made up of narrowly tailored conditional rules, essentially a specific rule for more or less every combination of circumstances that might arise, with each rule directing agents to perform whatever specific action would be utility maximizing in the circumstances in which it applies. And it would, Brandt argues, "have the property of being that set, general conformity with which will maximize utility."[13]

Lyons disagrees with Brandt about whether this vast moral code would be extensionally equivalent to act utilitarianism, strictly speaking.[14] Considerations of space militate against trying to resolve this disagreement, but happily little hangs on this. There is another way of construing what it means to adopt a rule that circumvents this problem, and most contemporary rule utilitarians have construed it in this way. This is to explicate the notion of adopting a moral code not in terms of complying perfectly with that code but rather in terms of "accepting" or, as I will usually say, "internalizing" the code. Brandt explicitly favors this alternative, as does Hooker;[15] some rule utilitarians prior to Brandt (such as Mill, perhaps) may have favored it as well, albeit implicitly. An agent who internalizes a moral code has a psychological disposition of some sort that gives her a motive to obey its rules. Internalizing a moral code is commonly taken to involve being disposed to feel compunction prospectively when one considers violating its rules and to feel guilt after the fact when one knows that one has done so. Brandt, for example, says that adopting a moral code usually means having intrinsic motivation to obey it, feeling guilt when one violates the code oneself and disapproving of others when they do so, believing that acting in accordance with the code is important, esteeming others who are motivated to comply with the code to an unusual degree, using special terminology like 'morally ought' in connection with the code, and believing that these motivations, feelings of guilt, feelings of approval or esteem, etc. are justified.[16]

If to adopt a code is to internalize it, then ideal-code rule utilitarianism is almost certainly not extensionally equivalent with act utilitarianism. We just cannot internalize a moral code with an astronomical number of rules. The same is true of a moral code with a smaller number of exceedingly long and complex rules, e.g., rules containing multiple

exception clauses. And while we perhaps could internalize the moral code that comprises only the single rule "Maximize utility," it would almost certainly not be utility maximizing for us to do so. The fact that we often cannot know which of our available actions would maximize utility means that even if we always tried to obey this one-rule code, we would still often choose actions that fell short of maximizing utility. In addition, our frequent uncertainty about which of the actions open to us would be optimific would make it easy for us to engage in self-interested rationalization, convincing ourselves that the action that would have the best outcome for ourselves would also maximize aggregate utility. Moreover, it would be hard for us to predict how other people who had internalized this code would act, since even if we knew that they had internalized it we would still not always know what action would appear to them to be utility maximizing; this would make it difficult for us to coordinate our behavior with theirs and so to enjoy the benefits of social cooperation. Finally, if we interpret the notion of adopting a rule in terms of internalizing it, then it seems that our utility calculation must take into account the "teaching costs" involved in getting people to internalize the rule. Given this, the chances that this one-rule code would be the one whose general adoption would maximize utility are even slimmer. This rule can require great personal sacrifice for the sake of people with whom our sympathy or fellow-feeling is minimal or non-existent, and so there might be significant costs involved in getting people to internalize it (or at least in getting them to internalize it strongly enough to be willing to make those sacrifices).

For these reasons, the "general internalization utility" of a plurality of rules that offer more specific guidance about how we should behave would almost certainly be higher than that of the single rule that requires us to always maximize utility.[17] Since we can only internalize so many rules, however, there is a limit to just how specific the guidance they offer us could be; it could not be specific enough for them to direct us to perform the optimific action in every set of circumstances.[18] Hence the code whose general internalization would be optimific would both require some actions that do not maximize utility and forbid some actions that do. Simply put, a version of ideal-code rule utilitarianism formulated in terms of internalization rather than perfect compliance should not collapse into act utilitarianism.

The "general" part of the notion of general adoption is also open to varying interpretations. There are (at least) two different questions that we might ask about it. The first is about the breadth of the group in question. Collective rule utilitarians can give different answers to this question. Some might maintain that the same moral code is authoritative for all of humanity, for example, and say that this is the code whose general adoption by all of humanity would maximize utility. Another possibility is to narrow the group in question, so that it includes only the members of a given society or, to narrow it even further, only those individuals who are members of a given society during a particular period of time. Hooker and Parfit opt for the first of these possibilities, Brandt the second, and (arguably, at least) Mill the third.[19]

The other question is what percentage of the members of a given group must adopt a moral code before the code can be said to be generally adopted among them. An obvious possibility, one embraced by Parfit, is to say that a moral code is only generally adopted among the members of a group if each and every one of them adopts it.[20] However, Brandt points out that identifying generality with universality in this way is problematic, because the rules that it would be utility maximizing for everyone to adopt would not contain any rules specifying what is to be done with members of the group who do not adopt the rules. Since realistically it will never be the case that all of the members of any broad group do adopt the authoritative moral code, this is obviously a significant lacuna. It might appear that the authoritative moral code could still include rules about how to handle those who reject the authoritative code, even though these rules would never be acted on in a world in which everyone in the group in question adopted it. However, as long as we interpret the notion of adopting a moral code in terms of internalizing it, there is reason to think otherwise. In that case, as mentioned above, the utilitarian calculus that determines which moral code is authoritative needs to take the teaching costs involved in inculcating the code in the members of future generations into account. As long as there are any teaching costs associated with getting people to adopt "superfluous" rules that would never be acted upon, a code that included them would never be the code that it was utility maximizing for everyone to internalize.

A common way of responding to this problem is to say that what it means for a code to be generally adopted by some group is for it to be adopted by most, but not all, of the group's members. Brandt and Hooker, for instance, both equate the general adoption of a moral code with its adoption by 90 percent, or at least roughly 90 percent, of the group's members. The code whose adoption by 90 percent of a group would be utility maximizing would presumably need to include rules for dealing with members of the group who do not adopt the code. Of course, in the real world far fewer than 90 percent of the people in a given society (much less in all of humanity) may adopt the moral code whose adoption by 90 percent of them would maximize utility. Even if the authoritative code contains rules for dealing with a small minority of group members who reject its rules, we might still worry about whether it would offer appropriate guidance if the people who adopt the code were the slender minority. We might also wonder what assumptions we are to make about the 10 percent of the population who reject the authoritative code when we are calculating which moral code it would be best for 90 percent of the people to adopt.[21] Do they adopt some other code, and if so what? Or are they nihilists? This seems very pertinent to deciding what rules for dealing with them ought to be included.[22]

Other versions of rule utilitarianism

Despite the central place in the literature that it currently occupies, collective ideal-code

rule utilitarianism is not the only form that rule utilitarianism can take, and some other variants of the view are also worthy of mention. One of these might be called "collective actual-code rule utilitarianism." According to this view, the rules of the authoritative moral code for a given society must be "actual" rules, in the sense of belonging to that society's conventional morality. However, only actual rules that satisfy some utilitarian criterion are taken to belong to the authoritative code. Richard Miller, for instance, has recently formulated the view as holding that actions are morally wrong just if they are proscribed by "legitimate" ethical rules in the society in which they would be performed, where a rule is legitimate if and only if it is a de facto rule in the society in question, includes a qualification that allows it to be set aside in order to avoid disastrously bad consequences, and "promotes the happiness of people affected by the rule." Favoring a "satisficing" over a maximizing utilitarian criterion, Miller adds that "To promote happiness a rule does not have to be ideal. It merely has to make life better than it would be otherwise."[23] While Mill's utilitarianism is open to different readings, and indeed I have already mentioned him in the course of discussing ideal-code rule utilitarianism, Miller claims that actual-code rule utilitarianism is very much in its spirit.[24]

Another variant might be termed "individual ideal-code rule utilitarianism." This form of rule utilitarianism drops the assumption that the authoritative moral code for a given person must also be authoritative for the other members of some broad group to which she belongs; it says, instead, that the authoritative code is the one that it would be best from the utilitarian standpoint for *her* to adopt, given other people as they actually are. (Some philosophers build the notion of collectivity into their definitions of 'rule utilitarianism', which suggests that they would not regard this theory as a form of rule utilitarianism at all,[25] but I can find no principled reason for this.) In one of the very few discussions of this view, D. H. Hodgson gives the theory a generally sympathetic reception.[26] While he stops short of endorsing it, he argues that the consequences of someone's subscribing to this theory would at least be better than those of her adopting the act-utilitarian principle as a "personal rule."[27] (For Hodgson, a personal rule is roughly equivalent to a rule that a person has internalized.) Hodgson assumes that someone who subscribes to individual rule utilitarianism would adopt as personal rules the rules that she believes the theory says that she should adopt. On this basis, he concludes that an individual rule utilitarian would very probably have personal rules that approximate the rules of her society's conventional morality, and he takes this to have some desirable consequences. But Hodgson assumes that people who subscribe to act utilitarianism will have no personal rules beyond the act-utilitarian principle itself. He ignores the possibility that act utilitarianism might require us to adopt additional personal rules as part of our "decision procedure," most or all of which might be conventional moral rules.[28] Yet R. M. Hare famously contends that act utilitarianism requires exactly this.[29] Thus Hodgson has his thumb on the scales when comparing these theories.

Finally, there is the "primitive rule utilitarianism" discussed by Lyons. While no one

advocates primitive rule utilitarianism as a moral theory, it is worth mentioning if only to ensure that none of the other rule utilitarianisms examined herein are conflated with it. According to primitive rule utilitarianism, "An act is right if, and only if, it conforms to a set of rules *conformity to which in the case in question* would maximize utility."[30] This makes primitive rule utilitarianism quite different both from any version of collective rule utilitarianism and from the sort of individual rule utilitarianism that I have just discussed, where the agent's authoritative moral code is the one that it would be utility maximizing for her to internalize. Lyons presents primitive rule utilitarianism as the rule-utilitarian equivalent of "utilitarian generalization," the view according to which the moral standing of an action depends on the consequences of everyone in the same position doing the same. He argues that utilitarian generalization is extensionally equivalent to act utilitarianism if we take "the same position" to include all of the agent's circumstances that would make any difference to the outcome of her action, including how many similar acts of the same type have already been performed, and he maintains that utilitarians cannot consistently construe "the same position" in any other way.[31] Inasmuch as primitive rule utilitarianism is extensionally equivalent to utilitarian generalization, Lyons contends, it is extensionally equivalent to act utilitarianism, too. He suggests that we might think of primitive rule utilitarianism's authoritative code as comprising a vast number of very lengthy rules. Each rule would have the form "*A* is wrong except when *B* or *C* or . . . ," where *A* is some fairly general category of actions whose generalized utility is negative, and *B*, *C*, and so on describe quite specific sets of circumstances in which the generalized utility of acts of type *A* is positive. For instance, *A* might be lying, and *B* might refer to situations in which the lie in question is being told to a murderer inquiring about the whereabouts of her next intended victim. Because primitive rule utilitarianism equates adopting a set of rules with complying with it, rather than internalizing it, there are no limits on how many rules its authoritative code might contain or how specifically their excepting conditions are described (and hence how long the rules might grow). This means that the code could be fine-grained enough to echo act utilitarianism's verdict about the moral standing of any action that any agent could perform in any situation.[32]

Attractions of rule utilitarianism

In this section, I will discuss some of the various considerations that have been offered in support of rule utilitarianism, concentrating once again on the collective ideal-code version of the theory. In general, I will say little or nothing about why the various philosophers whose arguments I survey here favor utilitarianism over, say, Kantianism; that would be too large an undertaking. Instead, I will focus most of my attention on the reasons that they give for thinking that rule utilitarianism is the best development of the general utilitarian idea, and in particular on why they consider it more compelling than act utilitarianism.

One route that some rule utilitarians have taken is to argue that rule utilitarianism is superior to act utilitarianism on utilitarian grounds. John Harsanyi, for instance, argues that "other things being equal, a rule-utilitarian society would enjoy a *much higher* level of social utility than an act-utilitarian society would."[33] He offers a variety of reasons why this should be the case. One is that if people know that they live in a society whose moral norms allow promises to be broken whenever this would result in even a slight gain in utility then they will "have much less *incentive* to plan their future activities on the expectation that promises made to them would be kept . . . [and] to perform useful services for people on the mere basis of promised future rewards, without any immediate compensation, etc."[34] The optimific moral rules about promising would spell out much narrower circumstances in which promises might allowably be broken. Another problem with act utilitarianism, according to Harsanyi, is that in virtually every case it demands that an individual perform one particular action, namely the utility-maximizing one. This ignores, Harsanyi says, the "procedural" utility connected with freely choosing what to do from a range of morally acceptable alternatives, a utility that he takes rule utilitarianism to offer.[35] A third consideration that Harsanyi discusses concerns what he calls the "coordination effect." He claims that if a society is made up of act utilitarians then they will be unable to take advantage of opportunities to produce desirable outcomes via collective action. Harsanyi illustrates this with a series of cases in which a certain number of votes are needed for the passage of some "socially very desirable policy measure."[36] Given that there is some small cost to voting, he says, an act utilitarian will choose to vote only if she expects that the measure will pass if and only if she does so. Each individual agent who reasons this way "is very unlikely to vote," so in a society entirely populated with such individuals the measure is virtually certain not to pass. In contrast, Harsanyi says, the optimific rule for voters generally to adopt is one that would direct them to vote. Harsanyi's explanation of why a rule-utilitarian society would enjoy higher utility than an act-utilitarian society resembles Brandt's analysis of why it would not be utility maximizing for people to internalize only the single rule "Maximize utility." But at that point Brandt is essentially taking for granted that rule utilitarianism is justified and is discussing the contents of the authoritative moral code. Harsanyi, in contrast, is aiming to establish that rule utilitarianism is justified by showing that act utilitarianism is not.

Given this aim, Harsanyi's reasoning is problematic in much the same way as Hodgson's. The act-utilitarian society in his comparison is one in which people have all adopted act utilitarianism as a theory and have all internalized *only* the act-utilitarian principle. That is a crude picture of what a society of act utilitarians would look like. He fails to demonstrate that a society of rule utilitarians would enjoy a higher level of utility than a society of sophisticated act utilitarians who have internalized rules like those recommended by Hare. Harsanyi's type of argument might have force if it could be shown that a society of rule utilitarians who have all internalized the ideal code would

enjoy more utility than a society of sophisticated act utilitarians.[37] (It might seem unrealistic and so unfair to rule utilitarianism to assume for the purposes of this comparison that real human act utilitarians would discover the optimific rules for them to internalize. But it is equally unrealistic to assume that a society of rule utilitarians would successfully discover and internalize the theory's authoritative moral code.[38] Realistically, people who agree that rule utilitarianism is the best moral theory might sharply disagree about what rules make up its moral standard.)

Other rule utilitarians have argued for the theory on the basis of its allegedly closer fit with our "moral intuitions" or "considered moral judgments" than act utilitarianism. We find this line of argument in Harsanyi as well, when he contends that while act utilitarianism is inconsistent with the "common sense" moral claims that individuals have rights or that parents have special obligations to care for their own children, rule utilitarianism both is consistent with these claims and has the power to explain why they are true.[39] Hooker also offers an argument of this sort, and in fact his case for rule consequentialism rests almost entirely on its putative closeness of fit with our considered moral judgments. Indeed, Hooker's reason for (at least provisionally) adopting a theory of the good that assigns intrinsic value to certain things other than welfare is that he takes this to allow his rule consequentialism to conform even more closely to our considered moral judgments than does rule utilitarianism. Hooker claims that his theory fits widely shared considered moral judgments more closely than any other moral theory except for "Ross-style pluralism," but while he concedes that pluralism is not inferior to his rule consequentialism in this respect, he argues that we should still favor rule consequentialism over pluralism because of the greater extent to which it possesses certain "formal virtues" of moral theories (e.g., simplicity and theoretical unity). Hooker frames this line of argument in terms of the "reflective equilibrium" method of moral theory selection.[40] The specific moral intuitions with which Hooker suggests that his theory coheres more closely than does act utilitarianism or any act-consequentialist theory include the judgment that promises cannot justifiably be broken whenever this would produce a slightly better outcome than keeping them and the judgment that morality does not demand the level of personal sacrifice that act utilitarianism has been shown to require by Peter Singer and others.[41]

A very different argument for rule utilitarianism has been attributed to Mill. This argument has three premises. The first is that for an action to be wrong is for it to be one that an ideal conscience would condemn, i.e., one that a person possessed of an ideal conscience would feel guilty about violating. This claim amounts to an analysis of what it means to call an action wrong. The second premise is the empirical claim that we feel guilt when we violate some set of rules that we have internalized, so that the conscience might be conceived of as an "enforcer" of rules. Finally, the third premise is the normative claim that the notion of an ideal conscience should be cashed out in utilitarian terms, so that the authoritative code enforced by the ideal conscience is the one whose

enforcement would maximize utility. In Mill one also finds the implicit empirical assumption that contemporary members of the same society will by and large internalize the same moral code, which suggests that the authoritative moral code for a given person – the code that she would feel guilty about violating, if she had an ideal conscience – is the one whose general adoption among the members of her society at that point in history would be utility maximizing.[42]

A somewhat similar strategy for defending rule utilitarianism is used by Brandt, who also construes what it is for an action to be wrong in a way that essentially guarantees that he will settle on collective ideal-code rule utilitarianism. Brandt denies that terms like 'morally wrong' or 'morally obligatory' have at present any precise settled meaning, but he uses 'morally wrong' to mean "would be prohibited by any moral code which all fully rational persons would tend to support, in preference to all others or to none at all, for the society of the agent, if they expected to spend a lifetime there."[43] (Brandt offers this as more than a stipulative definition; he says that philosophers "should try to make it the descriptive meaning of the phrase," i.e., should try to cause this to become the settled meaning of 'morally wrong' in common usage.)[44] For Brandt, then, the choice between moral theories boils down to a choice between social moral codes; the question is what set of rules people should internalize, and Brandt argues that fully informed and fully rational people would make this choice on the basis of which code would "expectably maximize the public well-being," since vividly representing to themselves all of the facts that are relevant to the choice would "fumigate" their preferences to the point that only self-interest and empathy would be left as grounds for choice.[45] Brandt's argument for rule utilitarianism, like the one just discussed in connection with Mill, can be read as an argument for what R. M. Adams calls "conscience utilitarianism" that is transformed into an argument for rule utilitarianism by way of the claim that what the conscience does is to enforce rules.[46] Note that although act utilitarians like Hare may largely agree with rule utilitarians like Brandt about what rules people should internalize, they will have a quite different conception of what it is for an action to be wrong, e.g., simply that it is not the most choiceworthy action or the one that we have the most reason to do.

Finally, Parfit contends that the best reconstruction of Kant's moral theory lends support to rule consequentialism. "Kantians," he writes, "could claim":

> Everyone ought to follow the principles whose universal acceptance everyone could rationally choose, or will. There are some principles whose universal acceptance would make everything go best. Everyone could rationally will that everyone accepts these principles. These are the only principles whose universal acceptance everyone could rationally will. Therefore UARC: These are the principles that everyone ought to follow.[47]

"UARC" is Parfit's abbreviation for "universal-acceptance rule consequentialism." (He

notes that with a few modifications the argument above could be transformed into an argument for "universal-following rule consequentialism," but he does not attempt to determine which version of the argument or the theory we should prefer.)[48] Parfit calls attention to the fact that this argument has no consequentialist premises, and in the next section we will see one reason why this is significant.[49] After presenting the above argument and concluding that "Kantian Contractualism," when properly understood, is extensionally equivalent to rule consequentialism, Parfit goes on to argue that both of these theories are equivalent to a third, namely the moral theory that he takes to be the best reconstruction of "Scanlonian Contractualism." He maintains that the convergence of these three theories, which results in what he calls the "Triple Theory," should give us additional reason to accept each of them.[50] While Parfit presents a case for rule consequentialism generally rather than for rule utilitarianism specifically, a rule utilitarian might embrace Parfit's case and supplement it with a further argument for the welfarist claim that, as Parfit puts it, "things go best when they go in the way that would, on the whole, benefit people most by giving them the greatest total sum of benefits minus burdens."[51]

Objections to rule utilitarianism

In this final section I will canvass several objections to rule utilitarianism. As with my overview of arguments for rule utilitarianism, the focus will be on objections that target rule utilitarianism specifically (in its collective ideal-code form), as opposed to more general objections to utilitarianism that would apply equally well to, say, rule and act utilitarianism. I will consider three such objections.

The first of these I can deal with swiftly, because it has already been mentioned. It is the so-called "collapse" objection, according to which rule utilitarianism collapses into act utilitarianism in the sense of being extensionally equivalent to it. As we saw previously, whether rule utilitarianism does indeed collapse into act utilitarianism depends on precisely how the view is formulated. For instance, if we interpret the notion of adopting a rule in terms of internalizing it, as opposed to obeying it, then the danger of collapse appears to be averted.

A second objection might be labeled the "rubber duck" objection, after the title of Frances Howard-Snyder's paper "Rule Consequentialism Is a Rubber Duck."[52] Howard-Snyder argues that rule utilitarianism, and indeed rule consequentialism more generally, is like a rubber duck in that it does not genuinely fit its name; a rubber duck is not really a duck, and rule utilitarianism is not really a utilitarian theory. Consequentialism, according to Howard-Snyder, is by definition "agent neutral" inasmuch as the states of affairs that it tells every agent to produce can be described without making essential reference to the agent. Act utilitarianism is a paradigmatic example of consequentialism so understood, since it simply directs each person to produce states of affairs that contain maximal

amounts of utility. In contrast, according to Howard-Snyder, a rule-consequentialist theory like rule utilitarianism will direct agents to produce states of affairs in which, for instance, *her* promises are kept or *she* does not kill. Rule-consequentialist theories are therefore "agent-centered," according to Howard-Snyder, and this means that they should properly be classified as deontological rather than consequentialist. Utilitarian moral theories are a subset of consequentialist ones, so if the theory that is the subject of this chapter is not a consequentialist view, then it is not a utilitarian view either. In the interest of accuracy, we might need to call it something like "maximizing welfarist deontology" instead.

Hooker argues that Howard-Snyder's definition of 'consequentialism' is too narrow.[53] He also makes a more important point, however, one to which Howard-Snyder also calls attention. This is that her objection is only to the use of names like 'rule consequentialism' and 'rule utilitarianism', not to the substance of the theories that commonly bear these names. Rule utilitarianism might still offer the best account of morality even if we have to call it something else.

This brings us to the last and perhaps most threatening objection to rule utilitarianism, the "incoherence" objection. As the name reveals, the incoherence objection asserts that the rule utilitarian is guilty of a kind of inconsistency. It assumes that the rule utilitarian's reason for believing that the authoritative moral code is the code whose general adoption would maximize utility must be that she is committed to the idea that maximizing utility is a goal of overriding importance, i.e., that any argument for rule utilitarianism must include some statement to this effect as a premise. But if maximizing utility is in fact a goal of overriding importance, the objection asserts, then whenever we face a choice between obeying the authoritative moral code and performing an action that violates this code but that would yield more utility, we have more reason to do the latter than the former. In effect, then, the incoherence objection asserts that there is an inconsistency between the content of rule utilitarianism on the one hand and any possible argument for the theory on the other. Smart gives a classic statement of this objection:

> Moral rules, on the extreme [i.e., act] utilitarian view, are rules of thumb only, but they are not bad rules of thumb. But if we *do* come to the conclusion that we should break the rule and if we have weighed in the balance our own fallibility and liability to personal bias, what good reason remains for keeping the rule? I can understand "it is optimific" as a reason for action, but why should "it is a member of a class of actions which are usually optimific" or "it is a member of a class of actions which as a class are more optimific than any alternative general class" be a good reason? You might as well say that . . . the Australian team should be composed entirely of the Harvey family because this would be better than composing it entirely of some other family.[54]

One way to summarize the objection is to say that rule utilitarians insist that we should

abide by a set of rules even when the same considerations that recommended those rules in the first place count in favor of breaking them, and from this it should be clear why the objection is sometimes called the "rule worship" objection. This is a powerful objection against certain combinations of specific rule-utilitarian theories and the arguments for them.[55] However, not all arguments for rule-utilitarian theories will necessarily include a premise that says that maximizing utility is a goal of overriding importance. Recall that Hooker argues for rule consequentialism based on its close fit with our considered moral judgments and that Parfit argues for it on Kantian grounds. Neither argument is premised on what Hooker calls "an overarching commitment to maximize the good."[56] Analogous arguments could potentially be given for rule utilitarianism specifically, arguments not premised on an overarching commitment to maximize utility. And if a rule utilitarian's argument for her theory does not force her to embrace such a commitment, then she can tell us that sometimes we are obligated to choose actions that will not maximize utility without being guilty of any inconsistency whatsoever.

Notes

1. Lyons, *Forms and Limits of Utilitarianism*, p. 11.

2. Brandt, *Ethical Theory*, p. 253. Brandt also coined 'act utilitarianism' (*Ethical Theory*, p. 380).

3. Smart, "Extreme and Restricted Utilitarianism."

4. See, e.g., Sinnott-Armstrong, "Consequentialism."

5. See, e.g., Cocking and Oakley, "Indirect Consequentialism, Friendship, and the Problem of Alienation."

6. Berkeley, *Passive Obedience*, p. 11.

7. Berkeley, *Passive Obedience*, p. 13.

8. Berkeley, *Passive Obedience*, pp. 13–14.

9. Kant, quoted in Guyer, *Kant on Freedom, Law, and Happiness*, p. 94.

10. Hooker, *Ideal Code, Real World*, pp. 34–37 and pp. 59–65. Hooker does describe his ascriptions of intrinsic value to both virtue and equality as tentative.

11. Rawls, "Two Concepts of Rules," p. 20, n. 1.

12. Smart, "An Outline of a System of Utilitarian Ethics," pp. 10–12.

13. Brandt, "Toward a Credible Form of Utilitarianism," pp. 122–123.

14. See, e.g., Lyons, *Forms and Limits of Utilitarianism*, pp. 137–139.

15. Hooker, *Ideal Code, Real World*, p. 78.

16. Brandt, *A Theory of the Good and the Right*, pp. 164–176; see also Hooker, *Ideal Code, Real World*, p. 91.

17. See Hooker, *Ideal Code, Real World*, pp. 94–95; Brandt, *A Theory of the Good and the Right*, pp. 271–277.

18. Lyons, "Utility and Rights," p. 163, n. 9; Hooker, *Ideal Code, Real World*, p. 96.

19. Hooker, *Ideal Code, Real World*, pp. 85–88; Parfit, *On What Matters*, vol. I, p. 377; Brandt, *A Theory of the Good and the Right*, pp. 179–182; D. E. Miller, *J. S. Mill*, p. 99.

20. Parfit, *On What Matters*, vol. 1, pp. 377–378.

21. H. M. Smith, "Measuring the Consequences of Rules," pp. 428–432.

22. For some recent attempts to explore alternative formulations of collective ideal-code rule utilitarianism meant to address these questions, see Ridge, "Introducing Variable-Rate Rule-Utilitarianism"; Hooker and Fletcher, "Variable *Versus* Fixed-Rate Rule-Utilitarianism"; and H. M. Smith, "Measuring the Consequences of Rules."

23. R. B. Miller, "Actual Rule Utilitarianism," p. 22.

24. R. B. Miller, "Actual Rule Utilitarianism," especially pp. 6–8; see also Donner, "Mill's Moral and Political Philosophy," pp. 53–54. For criticism of this view, see Hodgson, *Consequences of Utilitarianism*, pp. 26–32. For an extended defense of a view similar to actual-code rule utilitarianism, see C. D. Johnson, *Moral Legislation*.

25. E.g., Mulgan, *The Demands of Consequentialism*, pp. 53–54.

26. See also D. E. Miller, "Mill, Rule Utilitarianism, and the Incoherence Objection"; and Kahn, "Rule Consequentialism and Scope," p. 633 and pp. 641–644.

27. Hodgson, *Consequences of Utilitarianism*, pp. 63–77.

28. Hodgson, *Consequences of Utilitarianism*, pp. 38–39 and p. 48.

29. Hare, *Moral Thinking*, pp. 25–64.

30. Lyons, *Forms and Limits of Utilitarianism*, p. 139.

31. Lyons, *Forms and Limits of Utilitarianism*, pp. 62–118.

32. Lyons, *Forms and Limits of Utilitarianism*, pp. 121–139.

33. Harsanyi, "A Preference-Based Theory of Well-Being and a Rule-Utilitarian Theory of Morality," p. 296.

34. Harsanyi, "Rule Utilitarianism and Decision Theory," p. 37. See also Brandt, *A Theory of the Good and the Right*, pp. 275–276.

35. Harsanyi, "Some Epistemological Advantages of a Rule Utilitarian Position in Ethics," pp. 392–393.

36. Harsanyi, "Rule Utilitarianism and Decision Theory," p. 38.

37. Jonathan Riley contends that it would ("Defending Rule Utilitarianism").

38. Lyons, *Forms and Limits of Utilitarianism*, pp. 157–158.

39. Harsanyi, "Some Epistemological Advantages of a Rule Utilitarian Position in Ethics," pp. 390–391.

40. Hooker, *Ideal Code, Real World*, pp. 4–31; see also Hooker, "Reflective Equilibrium and Rule Consequentialism." For criticism of Hooker's use of this method, see D. E. Miller, "Hooker's Use and Abuse of Reflective Equilibrium."

41. Hooker, *Ideal Code, Real World*, pp. 145–158.

42. For this interpretation of Mill see D. E. Miller, *J. S. Mill*, pp. 79–110; a very similar reading was developed much earlier by Lyons in a series of papers that started to appear in the 1970s, most notably "Mill's Theory of Morality."

43. Brandt, *A Theory of the Good and the Right*, p. 194.

44. Brandt, *A Theory of the Good and the Right*, p. 195.

45. Brandt, *Facts, Values, and Morality*, p. 58 and p. 240.

46. Adams, "Motive Utilitarianism," p. 470.

47. Parfit, *On What Matters*, vol. 1, p. 400.

48. Parfit, *On What Matters*, vol. 1, pp. 405–407.

49. Parfit, *On What Matters*, vol. 1, pp. 401–402.

50. Parfit, *On What Matters*, vol. 1, pp. 411–419.

51. Parfit, *On What Matters*, vol. 1, p. 373.

52. Howard-Snyder, "Rule Consequentialism Is a Rubber Duck."

53. Hooker, *Ideal Code, Real World*, pp. 108–111.

54. Smart, "Extreme and Restricted Utilitarianism," p. 353. See also Smart, "An Outline of a System of Utilitarian Ethics," p. 10, and Foot, "Utilitarianism and the Virtues," p. 198.

55. In "Mill, Rule Utilitarianism, and the Incoherence Objection" I develop the understanding of the objection presented here at greater length and argue that Mill's rule utilitarianism falls victim to it.

56. Hooker, *Ideal Code, Real World*, p. 101.

8 Global utilitarianism

Julia Driver

Global utilitarianism (or, more generally, global consequentialism) is the view that everything is subject to utilitarian evaluation. In the purest form of the view, 'everything' means literally *everything*, including eye color, shampoo, trees, and rocks as well as actions, motives, and intentions. Further, in the purest form of the view, this evaluation is in terms of 'right' and 'wrong'. Thus, there are "right" eye colors just as there are "right" actions. The extension of moral evaluation to things like eye color is very controversial, and one issue is how to limit, in a principled way, moral evaluation so as not to include eye color and other features that seem poor objects of moral evaluation. Another issue has to do with whether or not to limit the evaluation in question to 'right'. Both of these issues will be explored below. The appeal of some form of global consequentialism, however, is that it provides a very nice framework for understanding phenomena such as normative ambivalence, where there seem to be confused intuitions about how to evaluate a particular action or state of affairs. It may also provide a way for consequentialists to accommodate our intuitions about agent-relative reasons in cases of *blameless wrongdoing*, which arise for consequentialists when according to consequentialism the agent has acted wrongly yet we do not view her as blameworthy because, with respect to another object of evaluation besides "action" (perhaps character), we fully approve of her. These issues will be explored in more detail below. The primary motivation behind global consequentialism is to show that a fully worked-out consequentialist normative ethics will be immune to criticisms that the theory ignores character and ethical norms that are not impartial.

Derek Parfit was one of the earliest to develop the idea of a global version of consequentialism: "Consequentialism covers, not just acts and outcomes, but also desires, dispositions, beliefs, emotions, the colour of our eyes, the climate, and everything else."[1] It has since been expanded on by Philip Pettit and Michael Smith.[2] In addition to considering various ways global consequentialism can be developed, this chapter looks at the background leading up to its development.

The view is generally contrasted with other versions of utilitarianism or consequentialism. For example, direct act utilitarianism is the view that the right action is the action that maximizes well-being, or can be expected to maximize well-being. It is direct, since the action is evaluated in terms of its own consequences, not the consequences of something else (such as a set of rules, or a cluster of dispositions). It provides a standard for evaluation. Failure to live up to the standard is a violation of one's moral duty or moral obligation. However, this theory was attacked for being overly focused on *act* evaluation. This criticism tended to be leveled by virtue ethicists, who argued that utilitarianism failed to properly consider the significance of character in moral

philosophy. Global utilitarianism meets these challenges by articulating a theory which holds that not just actions, but all things (or, all things relevant to agency) are to be evaluated using the utilitarian standard.

Background

In the mid to late twentieth century there was a flurry of literature criticizing utilitarian moral philosophy. One line of criticism, popular among virtue ethicists, was that utilitarianism focused on act evaluation along the lines of right, duty, and obligation – to the exclusion of character evaluation. Kantian and other deontological approaches to evaluation came under attack for the same reason. To some extent, this line of criticism was instigated by Elizabeth Anscombe, whose 1958 article "Modern Moral Philosophy" was widely taken to be a call for a return to good, old-fashioned, Aristotelian virtue ethics. With respect to utilitarianism, the criticism goes, a commitment to impartial consideration of consequences as the criterion of right action is deleterious (i.e., "corrupt") and fails to consider relevant features of human flourishing that virtue ethicists such as Aristotle are, theoretically, in a stronger position to consider. These are the virtues. Aristotle famously held that virtues involve having, for example, the appropriate sorts of emotional responses to value. Nothing like this seemed to figure into classical utilitarianism – at least as the early virtue ethicists saw it.

After the initial flurry of enthusiasm, other scholars, such as Robert Louden, voiced a cautious skepticism of virtue ethics as a genuine, full-fledged alternative to other normative ethical theories. It may be an alternative, for example, that simply mirrors defects of the views it seems to attack. *Eliminating*, for example, "thin" notions of evaluation, such as 'right' and 'wrong', is pretty radical. Further, would a focus on character then be guilty of "excluding" action?[3] Writers such as Louden called for a kind of hodge-podge new theory that would resist reduction. For example, one might hold that right actions are the ones that maximize production of the good, and then also hold that virtues are traits of character that exhibit human excellence completely irrespective of production of good effects. This would be a combination of act consequentialism and a theory of moral virtue that was not at all consequentialist. A veneer of theoretical unity might be provided by appealing to our basic intuitions that (perhaps) action and character are to be understood and evaluated in completely different ways. But this is completely unnecessary. Utilitarianism has had, all along, the capacity to account for and incorporate virtue evaluation, as well as the evaluation of other ethically significant aspects of agency such as motives and intentions. It never needed to be all about action. The utilitarian "reductive" enterprise works for these other forms of evaluation as well. It is true that in the earlier part of the twentieth century such expansion of the theory was overlooked and people did focus on act evaluation. But that is largely an historical artifact of how our cultures have progressed.[4] The legal systems of most countries focus on actions – people are not punished for being bad people independently of performing morally bad actions.

Larger societies also make it more difficult to gather the information required for reliable virtue evaluation, at least of people one is not very familiar with.

A number of philosophers who were writing around the time that Parfit's *Reasons and Persons* appeared also seemed to be making room for other forms of consequentialist evaluation. Peter Railton convincingly argued that sophisticated consequentialism encouraged the development of good-producing *dispositions*, not simply right *action*. In some cases, indeed, these will conflict. In those cases we get what I have termed "normative ambivalence"; a stable evaluation, or a unitary evaluation, is hard to achieve because we are really thinking about two different things: the agent's action and the character the agent is expressing through the action. Railton uses the now-classic example of Juan and Linda to illustrate. As Railton presents them, Juan and Linda are a couple who are separated geographically and Juan is committed to consequentialism. He needs to decide whether to spend money to visit his wife or to spend the same money in charitable pursuits where it will help those in more severe immediate need and therefore do more good. Juan's rationale for visiting his wife appeals to what is important to human happiness – for example, having a circle that one is a part of, being part of a loving relationship, and so forth. But this does not exempt one from morality either, or subjecting one's life to moral scrutiny. For the sophisticated consequentialist, the consequentialist standard regulates one's development as a moral agent as well as one's actions and yet it does not dominate. For the sake of efficiency and for the sake of truly achieving important aims that might be undermined by conscious deliberation about achieving those aims, we cultivate *dispositions*, modes of thought, and tendencies of various sorts that are good-producing. Actions are not the sole focus of the theory.

Roger Crisp similarly developed a form of utilitarianism of the virtues, or UV: "An agent ought to live virtuously, consulting the BU criterion only on certain special occasions."[5] The BU (Biographical Utilitarianism) criterion is: "Any individual ought to live in such a way that the total amount of utility in the history of the world is brought as close as possible to the maximum."[6] On this view, it is the BU criterion which provides the standard of right action, but UV which tells us how to go about living our lives. Crisp's project marks another effort at detailing how the consequentialist can accommodate the view that character, as well as action, is an important component to a full account of moral evaluation.

Thus, there was a movement in the normative ethics literature away from simply evaluating actions, and toward advocating for the usefulness of a theory incorporating other objects of evaluation, or *evaluands*: not only dispositions or virtues, but features of agency that are crucial to explaining action, such as motives and intentions. Global consequentialism offered another route in accomplishing this expansion. Character traits, motives, intentions, etc., are appropriate subjects for evaluation. For example, we might hold that a person did the right thing, though from a bad motive, or that the person intended to do the right thing but was misinformed and so acted badly. These more nuanced ways of evaluating conduct and persons added depth to moral evaluation.

The view

Consider the following case, *Clare*, presented first by Derek Parfit:

> Clare could either give her child some benefit, or give much greater benefits to some unfortunate stranger. Because she loves her child, she benefits him rather than the stranger.[7]

As part of the background to considering this case, it is plausibly stipulated that the best set of motives for a person to have would include strong love for one's children. Assuming however that Clare is a consequentialist, it seems reasonable to hold that she judges her action, in which she fails to maximize the good, as the wrong action. Her moral failure is due to the fact that she wants to help her child more than she wants to provide greater help to the stranger. However, this preference is itself due to the fact that she loves her child, and loving her child is, morally speaking, a very good motive to have. It helps to make her a good mother. Her action is wrong, and yet the motives that gave rise to it seem morally good; indeed, they seem like the "right" ones for a mother to have. Here, Parfit argues, we have a case of blameless wrongdoing because while Clare did act wrongly, we do not blame her for acting wrongly since she did so out of a good motive. Conversely, there will also be cases of moral immorality, where a person acts rightly, but out of a bad motivational set. "She did the right thing, but for the wrong reason" is the sort of evaluation one might typically hear in these sorts of cases. Global consequentialism provides a structure to such cases: we evaluate more than just actions. Because of this, we will get split moral verdicts when the moral quality of actions fails to match the moral quality of motives.[8]

Further, the global consequentialist who allows for such split verdicts will be able to give some accounting of our intuitions that seem to favor agent relativity. One classic problem for consequentialist theories is that such theories are committed to impartial norms – not only are norms applied impartially, but their content is impartial. For example, one is not morally justified in treating a member of one group better than a member of another group simply in virtue of something like group membership. The happiness of everyone counts equally; no one is intrinsically more important than anyone else. But, of course, this does not seem to apply when we consider norms that are peculiar to our family relationships and friendships. We do not blame Clare for preferring *her* child, because we recognize that it is generally good for parents to prefer their own children.

After Parfit's initial discussion of global utilitarianism, other writers pursued it. Philip Pettit and Michael Smith wrote an influential article in which they defend the view that alternatives to global consequentialism within the consequentialist framework cannot be

defended. They contrast global consequentialism with what they call "local consequentialism." Local consequentialism privileges some category of evaluand – such as actions. For example, local act consequentialism takes act evaluation to be primary. Acts are evaluated the same as they are in global consequentialism, but then other things, such as motives and motive sets, are evaluated in terms of their promotion of right acts.[9] Other examples are rule consequentialism, which privileges rules and defines right action in terms of rules, and "conscience" utilitarianism, which, as they note, R. M. Adams characterizes as the view that an act is obligatory "if and only if it would be demanded of us by the most useful kind of conscience we could have."[10] These various forms of local consequentialism were often suggested as ways to deflect criticism that tended to focus on shortcomings of act consequentialism. So, for example, on a rule-consequentialist view one might hold that it is in fact not right to kill one innocent person to save five innocent people (even though, on the face of it, such an action maximizes the good), because such an act conflicts with a rule (such as "Do not kill innocent people") which is part of the set of optimal rules.

Pettit and Smith pursue an indirect argument in support of global consequentialism by attacking various versions of local consequentialism. There are so many possible variations that the strategy must be merely suggestive (as they acknowledge). They present three versions of local motive consequentialism and show how each is defective, suggesting a general critical argument. If we consider the crudest form of motive consequentialism it would be something like the following: "the right acts are those which are caused by right motives."[11] They rightly hold this to be absurd. To use a Kantian example, it is right for the shopkeeper to give correct change, even though his motive is not a particularly moral one – he simply wants to maintain a reputation for fairness. A more sophisticated version holds that "right acts are those which would have been caused by possession of the right motives, rather than those that are actually so caused."[12] Pettit and Smith argue that this seems lacking in consequentialist justification. On this variation, acting as though one is rightly motivated leads to right action. But this is puzzling, since a person trying to comply with this principle might well try to act as though he were rightly motivated, and that is not what an actually rightly motivated person would do. The third variation they consider holds that "right acts are those which would have been caused by the motives that it would be best for someone to try to inculcate."[13] This version also fails in lacking a clear consequentialist justification. The same pattern of objection can be raised, they argue, for any local form of consequentialism. Thus, global consequentialism is supported.

Global consequentialism was also recently articulated by Shelly Kagan.[14] Kagan argues that theories can be partly understood in terms of their *evaluative focal points*. So, for standard act consequentialism, the evaluative focal point is action. And for standard rule consequentialism the evaluative focal point is rules (and acts are evaluated in reference to rules). Kagan explicitly holds that distinct focal points ought to be

evaluated directly rather than indirectly and that "the most plausible version of consequentialism will be direct with regard to everything."[15]

There are considerable theoretical advantages to global consequentialism. It allows for a much more nuanced account of moral evaluation, which, in turn, helps to explain instances of *normative ambivalence* in evaluation. The idea, discussed above, is that there are evaluative situations in which modes of evaluation come apart, so a positive evaluation and a negative evaluation of the same thing are both, in their own way, appropriate. This is the case, for example, when one does the right thing for the wrong reason; or the wrong thing for the right reason; or when one performs the right action but in so doing reveals a vice; or in performing the wrong action, reveals a virtue. Consider a case in which the only way to kill a truly evil dictator (we assume that he richly deserves to die and also that if he is not killed he will go on to kill many thousands of innocent people) is to strangle him with one's shoelace (he is very careful and will not allow any standard weapons in his presence). Well, this may be the right thing to do, but it also requires a certain degree of ruthlessness, which might itself be a vice. Indeed, one would probably be worried about someone who could strangle even an evil dictator with a shoelace and not be very upset about it. Global consequentialism can account for this ambivalence because there are two quite distinct *evaluands* in play in these cases. In some important respects this is like blameless wrongdoing, or moral immorality, in that the phenomenon is diagnosed by pointing to the fact that we are focusing on different evaluands. But blameless wrongdoing involves cases where there really is not much in the way of ambivalence at all: of course Clare should favor her child. Those are cases, instead, where what we are trying to account for is agent relativity. In cases of normative ambivalence, the phenomenology of the judgment is quite different.

Criticisms

One issue with global consequentialism has to do with the oddity of referring to things like climate, eye color, and shampoo as "right." This is odd because these objects are not agents, and we tend to intuitively restrict moral evaluation to features relevant to agency. Certainly, an appeal to what seems intuitively important, morally speaking, would rule out moral evaluations of things (as opposed to the people who make the things, for example). Moral agents are sensitive to reasons; climates are not. Climates do not respond one way or another to suffering.

Bart Streumer has suggested that this straightforward global consequentialism is flawed in that it runs up against a highly intuitive principle, the principle of "'ought' implies 'can'." The example he uses is the following:

> Suppose that it is raining where I am at the moment, and that it would maximize the good if it were sunny instead. Global consequentialism will then claim:
> (3)Its being sunny is right.

. . . this claim implies:

(4)I ought to bring about that it is sunny.[16]

Since, clearly, I cannot bring it about that it be sunny, it is not the case that I ought to bring it about that it be sunny, and global consequentialism is false. Even a highly circumscribed version, called semi-global consequentialism, fails. Semi-global consequentialism holds the following: "Everything that maximizes the good and that agents can bring about is right." Streumer makes use of the *Clare* case to illustrate the point. Again, Clare is capable of loving her child; also, she is capable of benefitting the stranger. Streumer takes away from this that the semi-global consequentialist is committed to the claim that "It is right for Clare to both love her child and benefit the stranger," and this further implies that "Clare ought to both love her child and benefit the stranger." Yet, she cannot do both.[17]

Campbell Brown defends semi-global consequentialism by holding that Streumer has made use of an agglomeration principle that, arguably, ought to be abandoned. We can only go from "Clare ought to love her child" and "Clare ought to benefit the stranger" to "Clare ought to both love her child and benefit the stranger" if we also accept

> *Agglomeration:* If an agent ought to bring about P and this agent ought to bring about Q, then this agent ought to bring about both P and Q.[18]

Brown makes the charge that Streumer has begged the question against the semi-global consequentialist by assuming *Agglomeration*. The semi-global consequentialist can simply deny it, which puts Streumer in the position of having to give an independent argument for it. But, as Brown notes, *Agglomeration* has been attacked in other contexts – such as the literature on moral dilemmas. Moral dilemmas are situations in which the agent is required to do one thing, as well as another, and cannot do both. This flies in the face of *Agglomeration*, and it is not obviously wrong to say that we should give up *Agglomeration* in order to preserve the powerful intuition that moral dilemmas are at least *possible*.[19]

The larger issue may have to do with putting a great deal of weight on 'ought' implies 'can' – the principle is problematic due to numerous ambiguities. In fact, each word in the principle is ambiguous, and what exactly it commits a theorist to will depend on how these ambiguities are resolved. Any objection to global consequentialism based on 'ought' implies 'can' will need to resolve these issues, e.g., is the 'can' that 'ought' implies a logical, metaphysical, physical, or psychological one?

A more inclusive approach

Another possible way to approach global consequentialism is to hold that not only should

we not privilege evaluands; we should also not privilege modes of evaluation. We can evaluate using virtue terms, for example, and not just terms of 'right' and 'wrong'.[20] This strategy is reminiscent of the strategy earlier consequentialists used to resolve other problems with the theory in light of criticisms from virtue ethicists. On this view, a disposition that leads (systematically) to good effects might be described as the "right" one or it might be described simply as a "virtue."[21] Thus, we can evaluate different things – action and character – in terms of something other than rightness. The key to a view counting as consequentialist has to do with how the moral quality of whatever it is we are talking about is cashed out. If it is understood solely in terms of production of the good, then it is consequentialist. Just as there is nothing that limits evaluands to actions, there is nothing that limits moral quality or modes of evaluation to 'right'.

One problem raised for every form of global consequentialism is that it provides conflicting guidance. This is the negative aspect of its ability to account for normative ambivalence – there is no distinctive way to cut through that ambivalence. Brad Hooker views this problem as raising a genuine paradox for global consequentialism:

> Suppose, on the whole and in the long run, the best decision procedure for you to accept is one that leads you to do act *x* now. But suppose also that in fact the act with the best consequences in this situation is not *x* but *y*. So global consequentialism tells you to use the best possible decision procedure but also not to do the act picked out by this decision procedure. That seems paradoxical.[22]

But judging normative ambivalence to be *appropriate* rather than problematic goes a long way toward mitigating the air of paradox.

One should do what maximizes the good but also be the sort of person who fails to maximize the good when that failure is the result of dispositions it is important for people to have – dispositions that are crucial to human happiness, for example. This still allows us to evaluate actions that the person of good character does as wrong, even though they may well be virtuous. We evaluate actions and dispositions in distinct ways – as right and as virtuous respectively, for example. It is not at all surprising that it is possible that guidance might be mixed, because guidance about what to *do* is different from guidance about what to *be*.

Notes

1. Parfit, *Reasons and Persons*, p. 25.

2. Pettit and Smith, "Global Consequentialism."

3. Louden, "On Some Vices of Virtue Ethics."

4. See Schneewind, "The Misfortunes of Virtue."

5. Crisp, "Utilitarianism and the Life of Virtue," p. 154.

6. Crisp, "Utilitarianism and the Life of Virtue," p. 142.

7. Parfit, *Reasons and Persons*, p. 32.

8. Cf. Dancy, *Moral Reasons*, pp. 240–247.

9. Pettit and Smith, "Global Consequentialism," p. 122.

10. Adams, "Motive Utilitarianism," p. 479.

11. Pettit and Smith, "Global Consequentialism," p. 125.

12. Pettit and Smith, "Global Consequentialism," p. 127.

13. Pettit and Smith, "Global Consequentialism," p. 129.

14. Kagan, *Normative Ethics*, pp. 198 ff.

15. Kagan, "Evaluative Focal Points," p. 151.

16. Streumer, "Can Consequentialism Cover Everything?" p. 240.

17. Streumer, "Can Consequentialism Cover Everything?" pp. 241–243.

18. See C. Brown, "Blameless Wrongdoing and Agglomeration," p. 223.

19. C. Brown, "Blameless Wrongdoing and Agglomeration," pp. 224–225. In a footnote, Brown acknowledges David Brink's previous argument against Agglomeration (p. 225, n. 11).

20. See Driver, *Consequentialism*, pp. 146–147.

21. Driver, *Uneasy Virtue*.

22. Hooker, "Rule Consequentialism," section 5.

9 Objectivism, subjectivism, and prospectivism

Elinor Mason

Imagine that you are driving home, a route you have driven hundreds of times, and you are approaching an intersection. You are in a hurry, it is the middle of the night, and you know that there are rarely other cars at this intersection. If you speed through without stopping and there are no other cars, you will get home faster. What should you do? Should you stop? What is the *right* thing for you to do? Intuitions about the case are bound to vary a bit, but my guess is that most people will say that the right thing to do is stop. After all, even though the risk is very small, it is a risk of something very bad happening, and not just to you, but to someone else. If there is another car coming you could kill the people in the other car. So, morally, it seems that you should stop.

Now let us say that, although you cannot know it of course, there are no cars coming the other way. What do you think the right thing to do is now? The issue is not one about risk anymore, so much as perceived risk. From the point of view of the universe, there is no risk. No car is coming, so no one will be hurt if you speed through. But you don't know that, because nothing has changed from your point of view; you still think it possible, though very unlikely, that a car is coming. Now any disagreement is primarily about the nature of rightness. Some people think that the right thing to do is determined by what will in fact have the best consequences, and it does not matter that the agent cannot know the outcome of her action for sure. Thus, in this case the right thing to do is speed through, as this will have the best consequences. Others think that what determines rightness is not what will actually happen, but the point of view of the agent. The right thing to do is stop, even though there is no car coming, because from the agent's point of view there is sufficient risk of a car coming. In this chapter I will examine the debate about this issue.

The positions

The view known as "objective consequentialism" (that terminology is more common than "objective utilitarianism," though the distinction is not relevant for my purposes here) is relatively clear.

> *Objective Consequentialism:* The right action is the one that actually would have the best consequences.[1]

It is worth noting that the range of actions that is relevant is just the range of actions that is possible for the agent. An objective consequentialist would not say that the right thing for the agent in the above example to do would be to put the car into rocket mode and

fly over the intersection. On the other hand, as critics of objectivism have pointed out, it is not entirely straightforward to say what is "possible for an agent." There are many things that I can do in a simple sense, but cannot do intentionally, or reliably, and this seems to make a difference to whether or not I can be morally required to do those things. I can beat Karpov at chess in the sense that I can physically make the moves that would amount to beating him. However, if I sat down and tried to beat Karpov at chess I probably would not do it. I will come back to this issue in my discussion of the "'ought' implies 'can'" principle below.

So objective consequentialism makes rightness dependent on what will actually happen. Subjective and prospective consequentialism, by contrast, are both based on the idea that rightness is relative to the agent's point of view. This needs a bit more elaboration. On the one hand, we might mean that rightness is relative to what the agent believes and predicts, regardless of how the agent came to those views. Call this view, which I will refine further below, "subjective consequentialism."

> *Subjective Consequentialism:* The right action is the one that the agent believes is required by consequentialism.[2]

On the other hand, we might think that rightness is relative to the agent's point of view in the sense that it is relative to the agent's knowledge about the world and how things will turn out, but is nonetheless objective in that the right thing to do is what it would be rational to do. Let us call that view "prospective consequentialism."

> *Prospective Consequentialism:* The right action is the one that best balances risk and good consequences.[3]

This terminology is fairly standard now. Other terminology appears in the literature, but for simplicity, even when talking about a writer who uses a different terminology, I will stick to these terms.

Refining subjectivism

There are really two different versions of subjectivism in the literature: "pure subjectivism," as defended by H. A. Prichard and W. D. Ross,[4] and "theory-relative subjectivism," as defended more recently by James Hudson, Holly Smith, and Fred Feldman.[5] Subjective consequentialism is a form of theory-relative subjectivism, and I shall focus primarily on that, but it is worth clarifying what pure subjectivism says.

Pure subjectivism about obligation bases rightness on the agent's estimation of the moral facts as well as of the non-moral facts. So according to pure subjectivism, the right thing to do is what the agent thinks is morally required given what beliefs the agent has

about the world. There is no requirement that the agent be rational or even reasonable in her beliefs. She could believe that morality requires her to obey the bunnies in her head, but so long as she does what she sincerely thinks is right, she is acting rightly according to pure subjectivism. Ross and Prichard both anticipate an objection that has been widely raised against pure subjectivism. The objection is that it is absurd to make rightness depend on what the agent thinks is right: this makes the agent the final arbiter of what *is* right. So, as Michael Zimmerman points out, agents always know what they ought to do (they can find out simply by introspecting); agents who don't believe that they have any obligations don't have any, and those who have horrendous moral beliefs, like Hitler, are acting rightly just so long as they believe that they are acting rightly.[6]

We must be careful not to conflate rightness and goodness. Goodness is independent of agents – the option that has the most goodness has the most goodness regardless of what agents think about it. But rightness is a different sort of thing. Rightness is rightness-for-particular-agents, and so we should not worry that it varies with them. The objectivist assumes that there is a simple relationship between rightness and goodness, that what has the most goodness simply is right. But, arguably, we should take into account the position of the agent, and different agents will be in different positions. There is no reason why rightness should not be relative to agents in some ways.

However, it is unclear how far this line of argument takes us. It is one thing to say that rightness may depend on what it is *reasonable* for an agent to do in the circumstances. It is another to say that rightness depends on what the agent *takes to be reasonable*. Prichard and Ross are right to point out that rightness can be relative to the agent in some ways – this seems beyond question. But it is not at all clear that the concept of rightness can be stretched so that it is completely relative to the agent. Full subjectivism has no anchor in anything objective, and this is why Zimmerman's criticisms apply.

Prichard and Ross could, and probably would, embrace the somewhat counterintuitive consequences of full subjectivism. But then it might reasonably be argued that they are primarily concerned with the agent's *conscience*, and that they are not really talking about rightness at all. In short, we have a deep and abiding intuition that the right thing to do is independent of us, at least in some respects. Full subjectivism can capture some of our intuitions, but the price is very high.

A more common strategy is to argue that we need subjective principles of rightness as *supplements* to the objective principles of a moral theory. This is theory-relative subjectivism. As mentioned above, Hudson, Smith, and Feldman have all tried to work out accounts along these lines. The basic idea is that a moral theory is composed of some basic objective instructions, for example, "promote the good," and then secondary (subjective) principles which should be used when it is unclear how to follow the basic instructions. The secondary principles are supposed to be usable in all circumstances, and different writers have different accounts of how that is achieved.

What is crucial about theory-relative subjectivism is that it takes a particular moral

theory as given. You do what is subjectively right when you do what you believe is the right thing to do *according to* (e.g.) utilitarianism. So someone who thinks she ought to do what the bunnies in her head tell her is not acting rightly, however sincere her belief that that is the right thing to do. Hudson justifies this as follows:

> The purpose of a moral theory (subjective utilitarianism, for example) is to tell the agent how she should use whatever information she has available at the moment of decision . . . Any moral theory, in telling the agent what to do, will ignore the agent's possible commitment to other moral theories. And while it should be considered a defect in a theory that it issues instructions that the agent does not know how to follow, it is not similarly a defect if it issues instructions that the agent decides not to follow because she does not believe the theory.[7]

An agent who acts rightly according to subjective consequentialism must be trying to do what consequentialism requires, so she must first know what consequentialism requires, at least in principle. The view is subjective in that she could have crazy ideas about everything else and yet still count as acting rightly.

Refining prospectivism

Prospectivism is often defined with reference to an example that has appeared in the literature in various forms. Donald Regan, Frank Jackson, and others have presented versions of this example. Jackson's version is as follows:

> Jill is a physician who has to decide on the correct treatment for her patient, John, who has a minor but not trivial skin complaint. She has three drugs to choose from: drug A, drug B, and drug C. Careful consideration of the literature has led her to the following opinions. Drug A is very likely to relieve the condition but will not completely cure it. One of drugs B and C will completely cure the skin condition; the other though will kill the patient, and there is no way that she can tell which of the two is the perfect cure and which the killer drug. What should Jill do?[8]

I said that prospective consequentialism is the view that the right action is the one that best balances risk and good consequences. Risk here is obviously risk from the point of view of the agent. Prospectivism starts with the agent's point of view, and then tells her to balance risk and possible good outcomes. According to prospectivism, Jill ought to choose the safe drug: the one that she knows has an acceptable but not optimal result. This accords with our common-sense reaction to the example.

However, there are various complexities in understanding prospectivism. We need to know what we mean by "the agent's point of view." Different prospectivists give slightly different accounts. We might take "the agent's point of view" to refer to the beliefs she

actually has, even if those beliefs are unreasonable. Alternatively, we might take it to refer to the beliefs that she should have given the evidence that is *available* to her, where availability is an objective notion: a rational or reasonable agent would avail herself of this evidence.

Probability assignments, or "credences" as they are sometimes called, form part of an agent's belief set, and these are particularly important for prospectivists. Prospectivists have to choose between making rightness depend on the agent's actual probability assignments, and making it dependent on the probability assignments that it would be rational or reasonable for her to have.

Then there is value itself, and here there is a more complex choice. Should rightness depend on the agent's actual value system, the value system that it would be reasonable for her to have (which might vary over time and cultures), or the value system that is in fact correct? Further, there can be more or less objective accounts of what it is rational or reasonable for an agent to believe, and a thorough taxonomy of possible views here should recognize that variation too.

Most defenders of prospectivism explicitly think in terms of an objective relationship between the agent and her evidence. Rightness depends on what it would be *reasonable* to believe at the time of action.[9] Avoiding the "problem of craziness," as we might call it, is the main virtue of prospectivism over subjectivism. According to subjective consequentialism, an agent who believes that she will produce the best (risk-balanced) consequences by wearing crystals and chanting is acting rightly when she does those things, no matter how crazy her beliefs about the world are. By contrast, prospective consequentialism can say that she is not acting rightly because her beliefs are unreasonable.

I use the term 'reasonable' here rather than 'rational' because full rationality seems too high a standard. Most agents are not fully rational. Even objective standards of rightness aim to apply to real agents. In some, perhaps slightly obscure, sense it is realistic to require agents to be reasonable, but it is not realistic to require them to be fully rational. Further, fully rational agents would not get themselves into many of the situations where real agents find themselves having to make decisions, and so it is unclear that the notion of what a rational agent would think in these circumstances even makes sense.

Prospective consequentialists must also clarify what they intend regarding the agent's grasp of value. Jackson explicitly says that his account of prospective consequentialism uses the correct account of value, which, for Jackson, is what a person ought to desire.[10] But prospectivists might prefer to apply the "reasonableness" requirement to the agent's value system. After all, it is possible that the truth about morality is very obscure in some cases, and that a reasonable grasp of it is sufficient for an agent to count as acting rightly.

Henceforth in discussing prospectivism I will be referring to what we might call "moderate objective prospectivism." The idea is that prospectivism should not have a

standard that is too high (full rationality) and at the same time should avoid too low a standard (actual beliefs). Instead, prospectivism should hold that the standard of right action depends on what a *reasonable* agent would believe in the circumstances. A reasonable agent is not one with exceptional powers of rationality, but one who is rational enough, good enough at making probability estimates, and good enough at knowing what the values at play in a situation are. Moderate objective prospectivism defines rightness in terms of reasonable beliefs, reasonable probability estimates, and a reasonable understanding of value.

The crucial element in prospectivism is that it takes uncertainty, that is, reasonable uncertainty, into account. Prospective consequentialism does not tell the agent to aim for the best outcome, but for the outcome that best balances risk and possible good consequences. This is the point of Jackson's example: Jill should not risk a very bad outcome in order to have a shot at the best possible outcome. Rather, she should compromise, by going for reasonably good consequences without any risk.

A question arises about exactly how it would be rational or reasonable to deal with uncertainty. Jackson favors what he calls a "decision-theoretic" account, according to which the rational way to deal with uncertainty is to maximize expected value.[11] It is a common assumption among prospectivists that the best way to deal with risk is by using the notion of expected value. To maximize expected value you must first have an idea of what the value of each possible outcome would be. So in Jackson's example, let us stipulate that the value of a partial cure is 50, the value of a complete cure is 100, and the value of death is –1,000. We get the expected value of an uncertain option by multiplying the value of each possibility by the probability that it will come about, and adding together the results for each possible outcome. For drug A we have just one possible outcome: we know that it will partially cure the patient. So the expected value is 50. Drugs B and C have the same expected value as each other. We know that one possibility is a complete cure, but there is only a 50-percent chance of that. So we multiply 100 by 0.5 and get 50. But we must also take into account the 50-percent chance of death, and add that to our 50. So we multiply –1,000 by 0.5 and get –500. Adding the expected values of the two possibilities gives us the expected value of the option: –450. Thus we maximize expected value by prescribing drug A.

All this does is formalize what is intuitively obvious, that the possibility of a very bad outcome makes an option unattractive, even when there is also a possibility of a very good outcome. The relative badness of the bad option as well as the probability that it will come about determines whether it is worth the risk. We would probably risk some minor side effects for a complete cure; we might even risk a non-trivial side effect. But a sizeable risk of death is clearly imprudent.

Thinking in terms of expected value, as Zimmerman points out in his defense of prospectivism, may not always be the most rational or morally appropriate way to deal with risk.[12] Perhaps some very bad outcomes should be avoided even when the risk is

very small indeed, particularly when the bad outcome would fall disproportionately on some group – perhaps a historically disadvantaged group. Thus Zimmerman prefers to stay neutral on the question of how to deal with risk, and I will follow him in defining prospectivism so that it is neutral between maximizing expected value and other accounts of rational ways of dealing with risk.

The arguments

There is a sense in which these various accounts of rightness are not in competition. They are different ideas and we use them for different things. Sometimes we are interested in what actually would have happened, because we want to learn about what to do in the future. Sometimes we are interested in what the agent thought: we are interested in the agent's conscience, even if we think the agent was misguided. And sometimes we are interested in what it would have been rational for the agent to do.[13]

However, moral philosophers are very invested in the term 'right', and so each camp tries to argue that its conception of rightness has the best claim to the word and the notion. Various argumentative strategies have been used. In what follows I shall examine the various strategies and make some necessary clarifications along the way.

Action guidance

Jackson's example appears to show that objectivism must be false. Let us stipulate that, in fact, drug B will cure the condition. Then, according to the standard conception, it would be objectively right for Jill to prescribe B. Jackson's point is that this is irrelevant, because clearly she ought to prescribe drug A. She should not be trying to do what is objectively right, and she should not be trying to do what is most likely to be objectively right. Rather, she should do what she knows is objectively wrong: prescribe the safe drug even though she knows that this option is less than the best.

If we are convinced by Jackson's claim that Jill ought to prescribe the safe drug, we have a strong argument against objectivism. Objectivism tells us that it would be wrong to prescribe the safe drug, and so apparently must deny that Jill should prescribe it. The prospective view says that Jill ought to do what is most rational given the reasonable gaps in her knowledge. In this fairly straightforward case it is clearly rational to choose the safe option. That is the best balance of risk and good consequences. Going prospective solves the problem raised by Jackson's example. What we ought to do is prescribe the safe drug, and that is also what is right according to prospectivism.

On the one hand, the argument here is simply an appeal to an example: prospectivism gives the intuitively right answer. However, hard-headed objectivists remain unconvinced. For example, Fred Feldman, in his early work, describes a case with the same structure as Jackson's, but insists that the right thing to do is what is objectively

best.[14] So appeal to example may not be enough.

Jackson's example is often understood as demonstrating and vindicating a requirement that moral theories provide action guidance. "Do what is best" is not action guiding, whereas supposedly, "Do what is prospectively best" is action guiding, even for agents who lack full information. This is what Jackson himself says. He says that

> the fact that a course of action would have the best results is not in itself a guide to action, for a guide to action must in some appropriate sense be present to the agent's mind. We need, if you like, a story from the inside of an agent to be part of any theory which is properly a theory in ethics.[15]

This argument is not a good defense of prospectivism. Prospective consequentialism gives an agent an instruction along the lines of, "Do what it would be reasonable to do given what you know and what is at stake here." Sometimes this is enough. We can make good judgments about simple gambles: I know to bring an umbrella when the chance of rain is very high and getting wet would be very bad. However, life is often more complex than this. Our everyday ethical choices about, for example, what to buy in the supermarket, concern the details and implications of trade arrangements, the environmental and human costs of farming and manufacturing methods, and so on.

Of course, even the reasonable person would not have all the facts to hand, but reasonableness does require that comparisons are made between all the relevant considerations in a fairly rational way, and that probabilities are assessed fairly rationally. We are notoriously bad at even quite simple probability puzzles. Our primitive brains overestimate some sorts of risk (rare dramatic horrors) and underestimate the mundane everyday risks (smoking and drinking).[16] So if, as prospectivism maintains, rightness is based on what we *should* think about risk, rightness will often be inaccessible to us.

The demand for action guidance pushes us toward subjectivism. It seems that subjectivism is the only action guiding theory, the only one that can provide "a story from the inside," as Jackson puts it. But of course, a theory-relative subjectivism such as subjective consequentialism requires that the agent knows in principle what the relevant theory requires. It is possible that an agent, through no fault of her own, does not know what consequentialism requires. It seems that such an agent will be unable to derive action guidance from subjective consequentialism.

Praise and blame

Another issue that may help us decide between objectivism, prospectivism, and subjectivism is the issue of when praise and blame are appropriate. Objectivists have argued that prospectivism does not capture rightness, but merely praiseworthiness,[17] while prospectivists have argued that *because* prospective rightness seems more closely

tied to praiseworthiness than objective rightness, it must be the correct account of rightness.[18]

However, the fact that both objectivism and prospectivism can give the agent instructions that do not make sense to her shows that the right action according to these views can become detached from the agent's own efforts and will. It would be possible for an agent to do the right thing *by accident*. Or, to put it in terms that are familiar from another strand of moral philosophy, it becomes a matter of luck whether an agent does the right thing. If a doctor in Jill's position were to prescribe a drug that she believes has a 50-percent chance of killing her patient – perhaps the doctor is just in a flippant mood, or perhaps she has gone through a depression and is feeling nihilistic – we would be reluctant to praise her even if the drug cured the patient. The point is that the possibility of "moral luck" (that is, luck in whether or not you do the right thing) undermines the connection between acting rightly and praiseworthiness.

It might seem that subjectivism's account of right and wrong action is most closely aligned with praise- and blameworthiness. After all, as both Ross and Prichard argue in their defenses of subjectivism,[19] it seems that we are praiseworthy and blameworthy only for what is under our control.

However, the situation is more complicated than that. First, blameworthiness can be inherited from previously blameworthy actions. So far I have assumed that it is not Jill's fault that she does not know which drug is the cure and which is the killer. But imagine that Jill does not know which is which because she had a hangover and missed the crucial lecture in medical school. This does not alter our intuitions about what Jill ought to do. She ought to prescribe the safe drug, and it does not matter how her ignorance came about; she ought not to risk the patient's death. But seeing her ignorance as culpable changes our intuitions about whether she would be *praiseworthy* in doing so. I think we are less inclined to call an act praiseworthy when it is polluted by past wrongdoing.[20]

Second, we do not always restrict praise and blame to actions that are under an agent's control. In fact, we often praise and blame people for their traits. This takes us back to a dispute about responsibility and morality that is often characterized (rather crudely) in terms of a disagreement between Kant and Aristotle. The Kantian view is that we are only praise- or blameworthy for our efforts of will: a naturally misanthropic person who manages to be friendly is thus more praiseworthy than the naturally gregarious person who is friendly without effort. On the Aristotelian view, our traits are themselves praise- or blameworthy. Someone who is not even tempted to murder his or her colleagues is more praiseworthy than someone who is tempted but manages to resist. My point here is not to settle this debate; it is to show that the notion of praiseworthiness is not a simple one, and so not one that we can use to settle the debate about whether rightness is objective, prospective, or subjective.

‘Ought’ implies ‘can’

A line of argument that is very closely related both to thoughts about action guidance and to thoughts about praise and blame is that objective consequentialism somehow violates the ‘ought’ implies ‘can’ principle. The point of the principle is that the concept of morality entails that we cannot be morally obliged to do something that is impossible for us. As I said in providing an overview of the positions, objective consequentialism does not violate the principle in the obvious sense. Objective consequentialism’s account of rightness picks out the best option the agent has: the best action *of those that are possible for the agent*. The agent may not know which actions those are, and she may not know how to do all of them, but if they were not in fact possible in a simple sense, they would not be options.

However, we might think that more than mere possibility is required. The thought can be traced back to Prichard. Prichard sees a very strong connection between obligation and responsibility, and his argument for subjectivism is intertwined with an argument for the interconnectedness of these notions.

In “Duty and Ignorance of Fact,” Prichard begins by considering whether objectivism or subjectivism coheres better with our ordinary thought, and concludes that although both have difficulties, subjectivism does better. But Prichard’s main argument for subjectivism is that, once we reflect, we should all agree that an obligation can only be an obligation to set oneself to do something, not to actually do it.[21] Prichard’s point is that we can only be obligated to do things that are under our control, but what actually happens as a result of our efforts is not under our control. So Prichard appeals to a very rich conception of ‘ought’ implies ‘can’, and in particular, a very rich sense of ‘can’: for Prichard, you can do something in the relevant sense only if you know that what you are trying to do will actually happen when you try to do it.

Prichard goes on to consider whether this makes any difference to the debate between objectivism and subjectivism, and of course it does. Once we have accepted that obligations are simply obligations to set oneself to do something, the force of the objective view is lost. We don’t think that obligation is about achieving something at all, so why think that it is about achieving what would actually be best?[22]

More recently, Frances Howard-Snyder has also argued that objectivism involves a violation of the ‘ought’ implies ‘can’ principle. Her view is not as extreme as Prichard’s. She does not think that the only thing that we can do is set ourselves to do something, but she thinks that we have to know *how* to do something in order to count as being able to do it. For example, according to Howard-Snyder, the instruction “You ought to beat Karpov at chess” violates the ‘ought’ implies ‘can’ principle, because you do not know which moves will amount to beating Karpov at chess. The very same action described differently (“Move your king’s pawn to K4, your king’s knight to B3, etc.”) does not violate the ‘ought’ implies ‘can’ principle. As Howard-Snyder points out, there is an

appearance of paradox here: the same action is both possible and not-possible for the same agent.[23] I can move my king's pawn, etc., but can I "beat Karpov"? The paradox is not serious: actions can be described in different ways, and different descriptions of actions make a difference to our judgments about those actions. Oedipus did not want to marry his mother![24]

The hard-headed objectivist replies that there is no violation of the 'ought' implies 'can' principle. The agent's options are limited to what is possible for her. So she can beat Karpov at chess, and it is thus possible that that is the right thing to do. So what is there left to say against objectivism? One strategy is to try to show that the objectivist's account of 'can' has implausible results. Dale Miller and Eric Wiland both point out that if objective consequentialism is true we almost never act rightly.[25] In Wiland's version of the argument, he shows that the hard-headed objectivist is committed to saying that any literate agent is morally obliged to write down the cure for AIDS or the great American novel (because we can make those marks on paper). An objectivist may balk at this result, but equally, a hard-headed objectivist may simply bite the bullet and admit that objectivism has the result that we rarely act rightly.

The appeal to the 'ought' implies 'can' principle is an appeal to something just as complex and controversial as the original dispute among objectivists, prospectivists, and subjectivists. If we think that arguments about which sense of 'can' is relevant are arguments about what we can be held responsible for, then we are pushed toward subjectivism. But we could deny that; we could say that 'ought' is not connected with responsibility (and hence some strong sense of 'can') in that way, and that what you ought to do is what would in fact be best, even though that is not something that you can reasonably be held responsible for.

The primary notion

There is certainly ambiguity in our obligation terms. But there remains a question about which is the *central* or *primary* term. As Zimmerman puts it, there must be some "overall" moral obligation: "It is with overall moral obligation that the morally conscientious person is primarily concerned."[26] Zimmerman uses the notion of the "morally conscientious person" to argue that the prospective sense of rightness is primary. Zimmerman's claim, backed up with a detailed analysis of what prospectivism entails, is that prospectivism captures the sense in which the morally conscientious person wants to act rightly.

There is a problem here. If we think about what the morally conscientious person will actually *do*, it seems that we cannot say more than that she will do what is subjectively right: she will try as hard as possible to make a rational assessment of the situation and act according to her own assessment of what is right. Why think more should be included in the description than that? Why think, in particular, that a morally conscientious person

must actually get it right? Zimmerman's own argument seems to be that full subjectivism is absurd, and so we should be prospectivists. Subjectivism does seem absurd in some ways (though as I said, it remains the case that we are sometimes interested in whether someone acted out of good conscience even when we think their act unfortunate), but this does not, in itself, provide an argument for *defining* the morally conscientious person as someone who is reasonable in their beliefs, probability estimates, and account of value. All we can say about the morally conscientious person is that she will do what she thinks best.

We might use the notion of the morally conscientious person differently. We might ask not "What would the morally conscientious person do?", but "What would the morally conscientious person ideally hope to do?" This question seems closer to a "central" or "primary" concept of rightness in that it anchors rightness to the standards of morality. We have a fairly deep intuition that rightness is idealized to some extent. Doing the right thing must involve meeting an objective standard. But if the question we are asking is, "What would the morally conscientious person ideally hope to do?", then we have nothing that distinguishes prospectivism from objectivism. In one sense the morally conscientious person hopes to do what is actually best, in another sense she hopes to do what is rational. These considerations might lead one to worry that the notion of the morally conscientious person is really no help at all here.

However, this may be unfair to Zimmerman. What is certainly plausible is that whereas the objective sense of rightness captures bestness, and the subjective sense captures conscience, the prospective sense comes closest to capturing what we mean when we talk about 'ought'. Perhaps there is no more to say than that.

One way to elaborate on the notion of primacy is to examine how our concepts work in everyday contexts. If we have a clear common-sense intuition about a term, then perhaps that is the primary sense of the term. This is how Prichard and Ross argue in defending the subjective view. The driving example I began with is a modification of Prichard's example. Prichard argues that we take ourselves to have a duty to slow down when entering a main road and that the subsequent discovery that nothing was coming does not change our view; we still think we had an obligation to slow down.[27] Although Prichard is clearly right that we do sometimes take it that our obligation was to do the cautious thing, this is not always true. What if the costs of slowing down were relatively high? What if, for example, we risked the car stalling and never starting again? In that case we might say, quite naturally, "It turns out I should not have slowed down after all."

A more recent version of the concept use argument has been deployed by Peter Graham to defend objectivism. Graham claims that when we get more information about a choice situation, we take it that we are getting information about what our obligations were all along; we don't think that our obligations have changed as a result of the new information.[28] It is true that the way that we speak often implies this, as Graham points out. But we can make the same criticism of Graham as of Prichard: it is not hard to think

of other cases where the way that we speak favors the opposing account of rightness. In fact, Prichard considers a case where there is disagreement about the facts determining obligation, and Prichard's conclusion is the opposite of Graham's. Prichard considers trying to change someone's mind about his obligation, and says, "when our attempt to change his opinion is over, then, whether we have or have not succeeded, the question of whether he is bound to do the action will turn on the nature of *his opinion* about the facts."[29] Prichard goes on to point out that this is not inconsistent with thinking that we have a duty to try to stop this person from doing what he takes his duty to be. As Prichard says, "if this were not so, few would fight conscientiously for their country."[30]

Assessments

Subjectivism assessed

The first three of the arguments I have surveyed, which are designed to defend prospectivism over objectivism, seem to point toward subjectivism. They all strive to establish a firm connection between rightness and the agent's will. They all start from the idea, roughly, that if an act is right it must be the case that it was up to the agent, and that the agent did it knowingly and deliberately. However, only pure subjectivism can provide the result that acting rightly is entirely under the agent's control, and pure subjectivism is a very unpopular theory, for reasons advanced by Zimmerman and elaborated above: pure subjectivism implies that agents are infallible, that there are no unknown obligations, and that you can act rightly even if your values are appalling.[31]

Theory-relative subjectivism, such as subjective consequentialism, is not vulnerable to all of Zimmerman's worries. First, it is not the case that agents know what their obligations are simply by introspecting. An agent has to know what consequentialism says about rightness, at least in a very general sense. Zimmerman's third point is closely related: that if a moral deviant such as Hitler thought he was acting rightly, then he was, which seems absurd. Subjective consequentialism has a ready answer. Hitler's moral theory was debased, and so although we may have to say that Hitler was acting rightly according to his own moral theory, he was clearly acting wrongly according to subjective consequentialism.

It is less clear that subjective consequentialism has an answer to Zimmerman's second worry, that agents who do not believe they have obligations do not have any. It depends on why the agent believes that she has no obligations. If that belief is based on a mistake about the non-moral facts, then her mistake erases her obligation. This is because subjective consequentialism is subjectivist about the non-moral facts. If an agent does not see any relevant factual considerations, then there are no relevant facts.

It is not clear what subjective consequentialism should say when the agent's mistake

or ignorance is moral. Most people have a lot of uncertainty about what the right moral theory is, or at least about what the relevant moral considerations are in a particular circumstance. It might seem that this uncertainty should be built into the options and taken into account subjectively; it should not be another layer of uncertainty beyond subjective obligation. But if there are various possible subjective obligations relative to different moral theories, then it can be unclear which is the actual subjective obligation. This leaves no resources to distinguish between the different subjective obligations, and so we would not have a usable 'ought' after all.

The subjective consequentialist has various possible responses to this. The first is to argue, as Hudson does, that it is beyond the purview of a moral theory to provide answers to those who do not believe in that moral theory.[32] Another possible line is to argue that although the moral theory itself does not give any guidance, it is possible to come up with a usable 'ought' even when we are in the grip of uncertainty about morality itself. Ted Lockhart has produced a detailed account of what we should do when faced with normative uncertainty. Lockhart's conclusion is that we should maximize expected moral rightness. Lockhart argues, against Hudson, that it is possible to compare different accounts of value, and thus to hedge our bets between different moral theories.[33]

Objectivism assessed

Objectivism has the advantage of being clear. However, objectivists have to deal with the extremely counterintuitive answer that objectivism gives in Jackson's example. Several recent objectivists have suggested that sometimes we ought to do something we know is wrong so as to avoid the risk of doing something that is even more wrong.[34] It is at least conceivable, as I suggested earlier, that it is not the case that we ought always to do what is right. We have these various notions: bestness, rightness, and ought-to-be-doneness, as well as subjective notions. It would seem natural to keep rightness and ought-to-be-doneness together; the most natural idea is surely that if something is right you ought to do it. However, it may be that our ideas have outstripped our terms, and that it would be better to separate ought-to-be-doneness and rightness, as these objectivists propose. If this is the case, we had better be very clear about what our revisionist meanings are.

If rightness and wrongness come in degrees and are not determinants of ought-to-be-doneness, as objectivists hold, we need to know how to rank options in terms of degrees of rightness and wrongness. Imagine that an agent has a choice between one option, A, which involves telling a lie, and two more, B and C, of which one will result in someone's death and the other will have only good results. She does not know which of B and C will cause a death. Let us assume that the objectivist will say that although lying is wrong, it is not as wrong as causing a death. So what determines this ranking? It is possible that the ranking is simply given, and that there is no explanation for it. But surely the objectivist will say that although lying is *bad*, it is not as bad as causing death, and so

what she should do here is lie. In other words, rankings that could be put in terms of less wrongness could equally be put in terms of less badness. (Both consequentialists and moderate deontologists – who allow some place for goodness as well as rightness in their theory – can say this.) In fact, it is hard to see any other way that the ranking could be justified.

In that case, we have the same answer that prospectivism gives (you ought to do what is prospectively best given the values at stake), but a different terminology. Whereas the prospectivist says that the prospectively best action ought to be done and is right, the objectivist suggestion being considered here is that the prospectively best action ought to be done even though it is wrong. Does anything really hang on saying that the prospectively best action is wrong rather than, what we all admit, not the best possible action in the circumstances?

I suspect that the disagreement between the prospectivist and the objectivist has evaporated here, and the only battle remaining is over the terminology. As for who wins that battle, it is clear that the prospectivist has the more intuitive use of the terms. We usually think of rightness as being tied to ought-to-be-doneness, and wrongness to ought-to-be-avoidedness. If rightness is to be separated from ought-to-be-doneness we need a good reason. It would have to be the case that rightness was indicating something that is important, and not covered by the other terms we have. But we have the term 'bestness'. Both consequentialists and moderate deontologists can say everything that needs to be said about the option that would in fact be best just by saying that it would be best.[35]

Prospectivism assessed

As I have argued above, prospectivism is not guaranteed to be action guiding, and it does not necessarily do any better than objectivism at cohering with our notions of praise and blame or at obeying the 'ought' implies 'can' principle. Thus many of the traditional arguments for prospectivism fail. There remain, however, two promising arguments for prospectivism. The first is, of course, that it seems to give the right answer in the Jackson case. The second, broader argument is that the prospective sense of rightness is the primary, central, or main sense of rightness. As I have shown in the previous section, the objectivist argument that we should sometimes act wrongly so as to avoid a greater wrong seems to rely on an unnecessarily clumsy use of the terminology. Insofar as prospectivism involves a more straightforward use of the terminology, prospective rightness may have a better claim to primacy.

Conclusion

It is clear that we use and understand objective, subjective, and prospective senses of our obligation terms. We may never eliminate the ambiguity in our uses of 'right' and 'ought'.

However, we have certainly made progress in understanding the distinctions between these different terms. We might see the various terms on a spectrum, ranging from fully objective, what would actually be best, through various degrees of prospectivism, to full subjectivism. Or, we might see the different notions as capturing quite different ideas – ideas about bestness (objectivism), ideas about reasonableness and evidence (prospectivism), and ideas about conscience and responsibility (subjectivism). So far, no clear winner has emerged from the philosophical disagreement about which of these senses is primary, but the debate is ongoing.

Parts of this chapter are based on my "Objectivism and Prospectivism about Rightness."

Notes

1. Defenders of objectivism about the right thing to do (not necessarily consequentialist) include Sidgwick, *The Methods of Ethics*; G. E. Moore, *Ethics*; W. D. Ross, *The Right and the Good*; Lyons, *Forms and Limits of Utilitarianism*; Bergström, *The Alternatives and Consequences of Actions*; Feldman, *Doing the Best We Can* and "Actual Utility"; Driver, *Consequentialism* and "What the Objective Standard is Good For"; and Graham, "In Defense of Objectivism about Moral Obligation."

2. The only prominent defenders of pure subjectivism about rightness are Prichard, *Duty and Ignorance of Fact*; and W. D. Ross, *Foundations of Ethics*. Neither is a consequentialist. Recent defenders of a diluted subjectivism include Hudson, "Subjectivization in Ethics"; H. M. Smith, "Subjective Rightness"; Feldman, "True and Useful"; and Mason, *Subjective Consequentialism*. All except Smith are broadly consequentialist.

3. Defenders of prospectivism (all more or less consequentialist except Ewing) include J. S. Mill, *Utilitarianism* (*Collected Works*, vol. x); Ewing, *The Definition of Good*; Smart, "An Outline of a System of Utilitarian Ethics"; Gruzalski, "Forseeable Consequence Utilitarianism"; Ellis, "Retrospective and Prospective Utilitarianism"; Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection"; Oddie and Menzies, "An Objectivist's Guide to Subjective Value"; Timmons, *Moral Theory*; Howard-Snyder, "The Rejection of Objective Consequentialism"; Zimmerman, "Is Moral Obligation Objective or Subjective?" and *Living with Uncertainty*; and Mason, "Objectivism and Prospectivism about Rightness."

4. See Prichard, "Duty and Ignorance of Fact"; and W. D. Ross, *Foundations of Ethics*.

5. See Hudson, "Subjectivization in Ethics"; H. M. Smith, "Subjective Rightness," p. 72; and Feldman, "True and Useful." Smith's view is not consequentialist.

6. Zimmerman, *Living with Uncertainty*, pp. 13–14.

7. Hudson, "Subjectivization in Ethics," p. 224.

8. Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," pp. 462–463; see also Regan, *Utilitarianism and Co-operation*, pp. 264–265, n. 1; Feldman, *Doing the Best We Can*, pp. 46–47; and Parfit, *On What Matters*, p. 159.

9. See Ewing, *The Definition of Good*; Oddie and Menzies, "An Objectivist's Guide to Subjective Value"; Gruzalski, "Forseeable Consequence Utilitarianism"; and Zimmerman, *Living with Uncertainty*.

10. Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," p. 464.

11. Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," pp. 464–466.

12. Zimmerman, *Living with Uncertainty*, pp. 51–56.

13. Both Ewing, *The Definition of Good*, and Parfit, *On What Matters*, emphasize that there are different senses of 'ought'.

14. Feldman, *Doing the Best We Can*, pp. 46–47.

15. Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," pp. 466–467.

16. See Kahneman and Tversky, "Subjective Probability."

17. Moore, *Ethics*, pp. 80–83; and Driver, "What the Objective Standard is Good For."

18. Gruzalski, "Forseeable Consequence Utilitarianism"; and Ellis, "Retrospective and Prospective Utilitarianism."

19. See Prichard, "Duty and Ignorance of Fact"; and W. D. Ross, *Foundations of Ethics*.

20. For a discussion of this issue see H. M. Smith, "Culpable Ignorance."

21. Prichard, "Duty and Ignorance of Fact," pp. 95–98.

22. Prichard, "Duty and Ignorance of Fact," p. 98.

23. Howard-Snyder, "The Rejection of Objective Consequentialism," pp. 245–246.

24. See also Mason, "Consequentialism and the 'Ought Implies Can' Principle."

25. D. E. Miller, "Actual-Consequence Act Utilitarianism and the Best Possible Humans"; and Wiland, "Monkeys, Typewriters, and Objective Consequentialism."

26. Zimmerman, *Living with Uncertainty*, p. 2. Ewing (*The Definition of Good*, pp. 122–144) accepts that there are three valid senses of 'ought' (corresponding to what I call the objective, the subjective, and the prospective) but argues that prospectivism uses the primary sense of 'ought'.

27. Prichard, "Duty and Ignorance of Fact," p. 93.

28. Graham, "In Defense of Objectivism about Moral Obligation," pp. 91–92.

29. Prichard, "Duty and Ignorance of Fact," p. 94.

30. Prichard, "Duty and Ignorance of Fact," p. 94.

31. Zimmerman, *Living with Uncertainty*, pp. 13–14.

32. Hudson, "Subjectivization in Ethics," p. 224.

33. Lockhart, *Moral Uncertainty and Its Consequences*. See also Sepielli, "What to Do When You Don't Know What to Do"; and Bykvist, "How to do Wrong Knowingly and Get Away with It," pp. 35–36.

34. Graham, "In Defense of Objectivism about Moral Obligation." This strategy is also suggested by Bykvist, "How to do Wrong Knowingly and Get Away with It," pp. 35–38; Portmore, *Commonsense Consequentialism*, pp. 15–16; and Driver, *Consequentialism*, p. 125. Smart says that what is right is what would actually produce best consequences ("An Outline of a System of Utilitarian Ethics," p. 47) and yet that one ought to do what would maximize probable benefit (p. 12), but he does not defend the divergence in concepts here.

35. For more on this see Mason, "Objectivism and Prospectivism about Rightness," pp. 17–19.

10 Subjective theories of well-being

Chris Heathwood

The topic of well-being

Classical hedonistic utilitarianism makes the following claims: that our fundamental moral obligation is to make the world as good as we can make it (consequentialism); that the world is made better just when the creatures in it are made better off (welfarism); and that creatures are made better off just in case they receive a greater balance of pleasure over pain (hedonism). The third of these claims is essentially a theory of well-being. Other forms of utilitarianism make use of different accounts of well-being, but whatever the version of utilitarianism, well-being appears in the foundations. Thus a complete examination of utilitarianism includes a study of well-being.

We can get at our topic in more familiar ways as well, and our topic is of interest independently of the role it plays in utilitarian theory. We can get at our topic by taking note of some obvious facts: that some lives go better than others; that some things that befall us in life are good, and others bad; that certain things are harmful to people and others beneficial. Each of these facts involves the concept of well-being, or welfare, or of a life going well for the person living it. Many other familiar expressions – 'quality of life', 'a life worth living', 'the good life', 'in one's best interest', 'What's in it for me?' – involve the same notion. We thus make claims about well-being all the time. Such claims naturally give rise to a philosophical question: What is it that makes a life go well or badly for the person living it?

Our question is not the perhaps more familiar question: What sorts of things *tend to cause* people to be better or worse off? It is interesting to investigate whether people's lives are made better by, say, winning the lottery, spending less time on the internet, or having children. But these are not the sorts of questions that philosophers of well-being ask. If your life would be made better by winning the lottery, this is due to the effects that winning the lottery would have on other features of your life, such as on your ability to pay for college or on the sorts of vacations you could take (and the value of these latter things might similarly lie wholly in their effects). But in the philosophy of well-being, we are trying to figure out what things are *in themselves* in our interest to have. We are asking, that is, what things are *intrinsically* good or bad for people, as opposed to what things are merely *instrumentally* good or bad for people.

Nor is our question: What things make *the world* intrinsically better or worse? The philosophical question of welfare is the question of what things are intrinsically good *for people*, and other subjects of welfare. But we also make claims about what things are good *period*, or good "from the point of view of the Universe."[1] For example, some

people believe that it is good in itself when something beautiful exists, even when no one will ever observe it. Whether or not this view is correct, philosophers of well-being are not asking about this kind of value. But it is easy to confuse it with well-being, because the clearest example of something that makes the world better is someone's having things go better for him or her. The claim that it is good when things go well for someone is not trivial, however. The easiest way to see this is to notice that it may have exceptions. It may fail to be a good thing, for example, when wicked people are well-off; perhaps it would be better if they were badly off.

Finally, our question is not: What sort of life makes for a *morally* good life? It seems that we can easily imagine someone leading a morally upstanding life that turns out to be of no benefit to her. But even if we became persuaded, through philosophical argument, that this is not possible, perhaps because moral virtue is its own reward, it still seems that being well-off and being moral are distinct phenomena.

It hardly needs arguing that the question of what makes a person's life go well is important. First, the question is just inherently interesting, and worth studying in its own right, even if answering it were relevant to no other important questions. It also has obvious practical implications: most of us want to get a good life, and knowing what one is might help us get one. Aside from these direct reasons to be interested, our topic is relevant to many of the most important questions we as people face. Most obviously, it is relevant to our moral obligations. This is of course true if utilitarianism is true, but it is no less true otherwise. For on any plausible moral theory, the effects that an act would have on the welfare of people and other animals is at least one morally relevant consideration. Utilitarianism stands out in claiming that well-being is the *only* basic morally relevant factor. Well-being also matters for politics. When deciding which political systems, institutions, and laws we ought to adopt, one obviously relevant factor is how well people will fare under the possible schemes. Well-being relates also to justice. One kind of justice, for instance, involves distributing welfare according to desert. The concept of well-being is also tied up with many virtues and vices, moral and non-moral. For example, a considerate person is one who frequently considers the interests of others, while a selfish person does this insufficiently. A person who can delay gratification for the sake of her long-term interests is a prudent person (this is why 'prudential value' is yet another synonym for 'well-being'). Welfare is probably also conceptually connected to each of the following phenomena: love, empathy, care, envy, pity, dread, reward, punishment, compassion, hatred, and malice. Seeing the connections that the concept of welfare has to other concepts can even help us to identify the very concept we mean to be asking about in the first place.

Subjective vs. objective theories of well-being

The distinction

One way to begin answering the question of what makes a person's life go well for him or her is simply to produce a list of things whose presence in our lives seems to make them better. Here is an incomplete list of some possibilities:

enjoyment
freedom
happiness
being respected
knowledge
health
achieving one's goals
friendship
getting what one wants
being a good person
being in love
creative activity
contemplating important questions
aesthetic appreciation
excelling at worthwhile activities

Most or all of these have opposites that are intuitively bad, but to keep things simpler, we will focus on the good things.

Something interesting about our list above is that all of the items on it are things that most people *enjoy*, and *want* in their lives. They are things we have positive attitudes toward (or, in some cases, they *just are* positive attitudes). This raises a question that is among the deepest and most central to the philosophical study of well-being: Are the things on the list above good solely in virtue of the positive attitudes that we have toward them, or do they benefit us whether or not we have these attitudes toward them? As Socrates might have put the question: Do we want these things in our lives because it is good to have them, or is it good to have them in our lives because we want them?[2] This is essentially the question of whether well-being is objective or subjective. Subjectivists maintain that something can benefit a person only if he wants it, likes it, or cares about it, or it otherwise connects up in some important way with some positive attitude of his. Objectivists deny this, holding that at least some of the things that make our lives better do so independently of our particular interests, likes, and cares.

What do we mean by 'positive attitude'? We mean to include attitudes of favoring something, wanting it, caring about it, valuing it, believing it valuable, liking it, trying to get it, having it as a goal, being fond of it, being for it, having an interest in it, and the like. Philosophers call these 'pro-attitudes'.[3] Not all subjective theories of well-being hold that all the attitudes just listed are relevant to well-being. A particular subjective theory will often single out one of them as *the* pro-attitude that is required for a person to be

benefitted.

In the next section, we will survey some of the particular varieties of subjective theory; in the remainder of this section, we will look at what is perhaps the most important reason for preferring the general subjective approach as well as a central reason for preferring an objective theory. In the process of doing this, we will further clarify the distinction between subjective and objective theories of well-being.

General considerations in support of subjectivism

Perhaps the main reason to think that the subjective approach is right is that there is a strong, widely shared intuition that suggests that the subjective approach is correct. This intuition is expressed in a frequently quoted passage by the philosopher Peter Railton:

> It does seem to me to capture an important feature of the concept of intrinsic value to say that what is intrinsically valuable for a person must have a connection with what he would find in some degree 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.[4]

Many share Railton's intuition. If we do, and if our evaluative intuitions are a guide to the truth about value, then this gives us reason to think that the subjective approach to well-being is the correct one. For Railton's intuition seems to be more or less just another way of putting the subjective approach.

If this sounds question-begging against the objectivist, a related way for the subjectivist to support her view is to elicit a similar intuition, but about a particular case. This might seem less question-begging. Here is such a case:

> Henry reads a philosophy book that makes an impression on him. The author defends an objective theory of well-being that includes many of the items on our sample list above. Henry wants to get a good life, and so he goes about trying to acquire these things. For example, to increase his knowledge – one of the basic, intrinsic goods of life, according to the author – Henry reads a textbook on entomology and acquires a vast knowledge of insects. Henry finds, however, that this new knowledge, as he puts it, "does nothing for me." He pursued it only because the author recommended it, and he can muster no enthusiasm for what he has learned, or for the fact that he has learned it. He in no way cares that he has all this new knowledge, and he never will care. It has no practical application to anything in his life, and it never will.

Now ask yourself: Was Henry benefitted by gaining this vast knowledge of entomology? The subjectivist expects that your judgment will be that, no, Henry was not benefitted. If

so, this supports subjectivism over objectivism about well-being. For objectivists who affirm the intrinsic value of knowledge are committed to saying that Henry was in fact benefitted by gaining this knowledge.

Objectivists who do not include knowledge on their list avoid this particular counterexample, but they will postulate other intrinsic goods, such as, say, freedom. The subjectivist will then ask us to imagine a new case: a case of someone who dutifully increases her share of the putative good – perhaps she moves to a state with fewer laws restricting her freedom – but who finds that she just does not care about having this new alleged good, and that it does not get her anything else that she cares about, wants, or likes. Because the putative good in question is objective – i.e., it bears no necessary connection to positive attitudes on the part of a subject who has it – it will always be possible for it to leave some people cold. If we share the intuition that such people receive no benefit when they receive the alleged good, we have a counterexample to the objective theory in question.

Some putative goods on the list above are not objective. Consider happiness, or at least one kind of happiness: being happy about something in your life, such as your job. Being happy about your job does bear a necessary connection to a positive attitude of yours, because being happy about your job *is* one such attitude. Being happy about your job *cannot* leave you cold, since the very attitude of being happy about your job is an attitude of finding something to some degree compelling or attractive. Thus we cannot construct a case analogous to the case of Henry about the putative good of being happy. This will not help objectivists, of course, since a theory that claims that the single, fundamental human good is *being happy* is a subjective rather than an objective theory.

Other putative goods on the list above are clearly objective. Knowledge, if an intrinsic welfare good, is an objective one because it need not connect up in any way with our pro-attitudes. Note that this is true even though knowledge is (at least in part) a mental state. Thus it is a mistake to understand the objective–subjective distinction as it is used in the philosophy of well-being as involving merely the distinction between states of the world and states of mind. To be a subjectivist about well-being, it is not enough to hold that well-being is wholly determined by subjective states, or mental states. It has to be the right kind of subjective state – a "pro" or "con" mental state.

Further clarification of the distinction

It is worth making a further clarification about subjectivism. As we noted earlier, a Socratic way to think of subjectivism about well-being is as the view that things are good for people in virtue of the pro-attitudes they take toward those things. We also said that the theory that happiness is the good is a subjective theory. But consider someone who, while very happy about many things, never stops to consider her own happiness, and so never takes up any pro- or con-attitudes toward it. If the Socratic way of understanding subjectivism is literally correct, then the happiness theory will count as a form of

objectivism. For, as this example illustrates, it is possible on this theory for something (namely, being happy) to be good for someone without her taking up any pro-attitudes toward that thing.

One way to try to handle this is to reject the Socratic understanding of subjectivism as too narrow, and to hold that

> a theory is subjective just in case it implies the following: that something is intrinsically good for someone just in case either (i) she has a certain pro-attitude toward it, or (ii) it itself involves a certain pro-attitude of hers toward something.

This criterion counts the happiness theory as a subjective theory because, on the happiness theory, the only thing that is intrinsically good for people is a thing – their being happy about something – that itself involves their own pro-attitudes toward something (their being happy about something *just is* a pro-attitude toward something). This will be our official understanding of subjectivism about well-being. Correspondingly, objectivism about well-being is the view that at least one fundamental, intrinsic human good does not involve any pro-attitudes on the part of the subject.

General considerations in support of objectivism

One motivation for being an objectivist about well-being is that it just sounds plausible to say that things like freedom, respect, knowledge, health, and love make our lives better. But we have to be careful. Subjectivists can agree with this plausible thought, since they know that most people have pro-attitudes toward these things, or at least that these things cause most people to have pro-attitudes (such as happiness or enjoyment) toward other things. Thus when these people get the things on the list above, their lives will be made better even according to subjectivism. To put it another way, subjectivists hold that the things on this list are typically *instrumentally* good for us to have, and hope to fully account for their intuitive value in this way.

However, some objectivists will continue to insist that the value of at least some such items is intrinsic and attitude-independent. In support of this, they might offer the following kind of argument against subjectivism. It begins by imagining someone who has bizarre interests, or, perhaps more effectively, base or immoral interests. Thus, John Rawls "imagine[s] someone whose only pleasure is to count blades of grass in . . . park squares and well-trimmed lawns."[5] G. E. Moore compares "the state of mind of a drunkard, when he is intensely pleased with breaking crockery" to "that of a man who is fully realising all that is exquisite in the tragedy of King Lear."[6] As an example of a morally corrupt interest, we can imagine a pedophile engaging in the immoral activities he very much wants to be engaging in. Finally, Thomas Nagel has us "[s]uppose an intelligent person receives a brain injury that reduces him to the mental condition of a contented infant, and that such desires as remain to him are satisfied by a custodian, so

that he is free from care." Nagel claims that "[s]uch a development would be widely regarded as a severe misfortune, not only for his friends and relations, or for society, but also, and primarily, for the person himself . . . He is the one we pity, though of course he does not mind his condition."[7]

According to the objection, subjective theories are committed to the following: that Rawls's grass-counter can get a great life by doing nothing more than counting blades of grass all day; that, so long as the amount of pleasure is the same between the two cases, it is just as well, in terms of how good it makes your life, to break crockery while drunk as it is to appreciate great art; that it is, at least considered in itself, a great good for the pedophile when he molests children; and that the brain-injury victim has in fact suffered no misfortune, so long as the desires that remain to him are well enough satisfied. But, the argument continues, surely claims such as these are implausible. One kind of evidence for this may be that we would not want someone we love, such as our own child, to live a life like any of the lives imagined here. We can avoid these putatively implausible claims by including objective elements into our theory of well-being, such as that exposure to great art is intrinsically good for people or that engaging in immoral activities is intrinsically bad for people.

To these objections, some subjectivists (including Rawls himself) "bite the bullet." They think that, on reflection, such lives in fact can be good *for the people living them*. After all, these activities are just the sorts of activities they want to be doing, and like doing. This may be easier to swallow when we remind ourselves that accepting such a claim does not commit one to the view that these lives are *morally good*, or that they manifest *excellence*, or that they are good in other ways that are distinct from their being beneficial to those living them.

One's ultimate view concerning such cases, and concerning the considerations above in support of subjectivism, will help determine where one stands on this most important philosophical question of well-being: whether to accept a subjective or an objective theory.

Before discussing specific kinds of subjective theory, it is worth mentioning a third option, one we will not have space to explore here: a hybrid of subjectivism and objectivism. According to *the hybrid theory*, well-being consists in receiving things that (1) the subject has some pro-attitude toward (or that otherwise involve pro-attitudes on the part of the subject) and that (2) have some value, or special status, independent of these attitudes. One's life goes better not simply when one gets what one wants or likes, but when one is wanting or liking, and getting, *the right things*. These might include some of the things on our list above. It is very much worth investigating the extent to which the arguments and considerations discussed in this essay apply to hybrid theories of well-being.[8]

Varieties of subjectivism

On one popular taxonomy, there are three main kinds of theory of well-being:

> *hedonism*, according to which pleasure or enjoyment is the only thing that ultimately makes a life worth living;
> *the desire theory*, according to which what is ultimately in a person's interest is getting what he wants, whatever it is; and
> *objectivism*, according to which at least some of what intrinsically makes our lives better does so whether or not we enjoy it or want it.[9]

We have already discussed objectivism (and it is discussed in greater depth in the next chapter). The desire theory is the paradigmatic version of the subjective approach to well-being. Hedonism is often also classified as a subjective theory, though, as we will see, this issue is somewhat complicated. In what remains, we will introduce and briefly explore hedonism, including how to classify it, and conclude with a lengthier treatment of the desire theory of welfare. Along the way, we will briefly discuss two kinds of subjective theory that may or may not be covered by the above taxonomy: *eudaimonism*, the view, often associated with hedonism, that well-being consists in happiness; and *the aim achievement theory*, the view, often associated with the desire theory, that successfully achieving our aims is what makes our lives go well. A related subjective theory, which we will not have space to discuss, appeals not to the subject's desires or aims, but to the subject's *values*.[10]

Hedonism

Hedonism is among the oldest of philosophical doctrines still discussed and defended today, dating back to the Indian philosopher Cārvāka around 600 BCE and the Greek philosopher Aristippus around 400 BCE.[11] The notions that suffering is bad for the one suffering and enjoyment good for the one getting it are intuitive raw data that any plausible theory of well-being must accommodate. Hedonism is controversial largely because it claims that *nothing else* is of fundamental intrinsic significance to how well our lives go.

In ordinary language, the term 'hedonist' connotes a decadent, self-indulgent devotion to the gratification of sensual and gastronomic desires. But it is no part of the philosophical doctrine of hedonism that this is the way to live. Hedonism is not the egoistic view that only one's own pleasures and pains should concern one, and hedonists often emphasize the greater reliability, permanence, and freedom from painful side effects of intellectual, aesthetic, and moral pleasures.

The most popular argument, historically speaking, for hedonism about well-being appeals to a theory of human motivation known as *psychological hedonism*.[12] According

to psychological hedonism, the only thing that anyone ever desires for its own sake is his own pleasure (ignoring pain here for brevity). Thus, whenever a person desires something other than his own pleasure, he desires it as a means to his own pleasure. The argument from psychological hedonism uses this psychological claim as a premise in establishing the conclusion that the only thing that is intrinsically good for someone is his own pleasure. To move from this premise to this conclusion, the argument requires the additional, often suppressed premise that only what a person desires for its own sake is intrinsically good for him.

This argument is almost universally rejected nowadays, even by hedonists.[13] Not only does the sweeping generalization of psychological hedonism seem too simplistic, the second premise – that only what a person desires for its own sake is intrinsically good for him – is evidently an abandonment of hedonism as the fundamental truth about well-being and a move to the desire theory.

This raises the question of just what relation pleasure has to our pro-attitudes, and this, in turn, bears on the question of whether hedonism should count as an objective or a subjective theory. There are two main views of the nature of pleasure. On the *felt-quality theory*, pleasure is a single, uniform sensation or feeling, in the same general category as itch sensations or nauseous feelings (only pleasant!). On the *attitudinal theory*, pleasure fundamentally is, or involves, an attitude – a pro-attitude that we can take up toward other mental states, like itches and nauseous feelings, or states of the world.

It would seem that whether hedonism qualifies as an objective or a subjective theory depends on which general approach to the nature of pleasure is correct. If a felt-quality theory is true, and pleasure is just one feeling among others, a feeling one may or may not care about, want, or like, then pleasure, if good, would seem to be an objective good, and hedonism an objective theory of well-being. But if pleasure is instead a pro-attitude, or essentially involves some pro-attitude, then pleasure, if good, would seem to be a subjective good, and hedonism a subjective theory of well-being. It is important to recognize that the issue here is not merely one of taxonomy. If hedonism is an objective theory, then it, like other objective theories, is committed to the perhaps counterintuitive idea that something some people may find in no way attractive, or that in no way connects to any positive attitudes of theirs, is nonetheless of benefit to them. If hedonism is a subjective theory, it avoids this implication.

If hedonism is a subjective theory, due to pleasure's being explainable in terms of some pro-attitude, does it remain a distinctive theory, or does it instead become a version of whatever kind of theory enshrines this pro-attitude? It depends upon which pro-attitude pleasure is ultimately explained in terms of. According to the contemporary hedonist Fred Feldman, all pleasure is ultimately explained in terms of the pro-attitude of *being pleased* that something is the case.[14] Since, according to Feldman, this hedonic pro-attitude cannot, in turn, be explained in terms of any other, non-hedonic attitude,

Feldman's theory of pleasure allows hedonism to remain a distinctive theory. Other attitudinal theories of pleasure reduce pleasure to other kinds of attitude, most commonly desire.[15] If the desire theory of pleasure is true, then a theory claiming that pleasure is the good might be best classified as a form of the desire theory of well-being.[16] The attitudinal theory of pleasure thus holds promise for hedonism – allowing it to avoid the problems of objectivism – as well as risk – the risk that it would cease to be a distinctive theory of well-being.

Interestingly, even if hedonism turns out to be an objective theory, it still faces some of the problems of subjectivism. For whatever pleasure turns out to be, people can get it from sources pointless, base, immoral, or unfitting (see the section "General considerations in support of objectivism," above). John Stuart Mill believes that hedonists can answer these objections by assigning value to pleasures on the basis not only of their intensity and duration, but their *quality* as well. He holds that intellectual, aesthetic, and moral pleasures, for example, have "a much higher value" than bodily pleasures of equal intensity and duration.[17] Thus a life full of the pleasures of studying Shakespeare could be a better life than one devoted to breaking crockery while drunk, even if the latter contains a greater *quantity* of pleasure. Mill bases his assessment of the greater value of these "higher" pleasures on the contention that people who are acquainted with both higher and lower pleasures invariably prefer the former (though this raises again the specter of hedonism's collapse into the desire theory). Some critics charge that Mill's appeal to quality is in fact an abandonment of hedonism.[18]

Hedonism, whether it turns out to be an objective or a subjective theory, also faces an objection that is more or less distinctive to it – avoided by competing theories, subjective and objective alike. The objection exploits the fact that pleasure is a mental state, and so that on hedonism, how well one's life goes is directly determined solely by one's mental states, and not by the way the external world is. The most vivid and well-known version of this objection is based on a thought experiment by Robert Nozick:

> Suppose there was an experience machine that would give you any experience you desired. Super-duper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain. Should you plug into this machine for life, preprogramming your life experiences? . . . [W]hile in the tank you won't know that you're there; you'll think that it's all actually happening . . . Would you plug in?[19]

If you would not plug in, and if one reason you would not plug in is that you think you would fare less well in such a life than in your actual, less pleasurable life, then this may show that you reject hedonism.[20]

Let us turn briefly to eudaimonism, the truistic-sounding theory that the good life is

the happy life. This is in fact no truism (assuming of course that by 'a happy life' we do not simply *mean* a good life). For it, too, faces the experience-machine objection as well as the problems of subjectivism more generally. In fact, given one leading theory of happiness – the view that to be happy is to have a favorable balance of pleasure over pain – hedonism and eudaimonism amount to the same view. According to a rival theory of happiness, to be happy is to be satisfied with one's life as a whole. This conception of happiness yields a different version of eudaimonism, one that comes apart from hedonism, since it is possible to be dissatisfied with a pleasant life.[21]

The desire theory

A person living her life on an experience machine will, unbeknownst to her, get little of what she wants. While some of our desires concern our experiences (we desire the taste of some food, or for some itch to be gone), many of our desires concern states of the external world (we desire to climb some mountain, or to be loved by someone). These latter kinds of desire will go unsatisfied on the experience machine (although the subject will often believe, falsely, that they are satisfied). According to the desire or preference-satisfaction theory of welfare, well-being consists in getting what one wants. On this theory, a life on the experience machine will be in many ways worse than an ordinary life, in which many desires about the external world are satisfied. The desire theory of welfare thus appears to avoid the experience-machine objection.

Perhaps the earliest discussion of the desire theory of any depth is found in Henry Sidgwick's *The Methods of Ethics*, though it may have been endorsed centuries earlier by Thomas Hobbes and also Baruch Spinoza.[22] It gained prominence in the twentieth century with the rise of welfare economics and decision theory, where preference theories of well-being or utility are often simply assumed. Economists may be motivated to assume the theory because it is thought to make well-being easier to measure than it would be on hedonism, since our desires are thought to be revealed through our choices, especially in free markets. Others have been motivated to accept the desire theory rather than an objective theory because they believe the former to fit better with a naturalistic worldview. Objective theories that posit more than one basic good also face a problem concerning how to compare goods of very different kinds. Monistic theories like the desire theory avoid this. Today the desire theory is often regarded as the leading theory of well-being, especially among utilitarians.[23]

Another putative attraction of the desire theory is that it very straightforwardly conforms to the intuition, introduced earlier, that what is intrinsically good for a person must be something he or she finds to some degree compelling or attractive. For to desire something, whatever else it is, is surely to find it to some degree compelling or attractive. As the theory that most clearly conforms to this intuition and as the theory that makes use of what is perhaps *the* fundamental pro-attitude, the desire theory is the paradigmatic subjective theory of well-being.

The simplest version of the desire theory of welfare claims that the satisfaction of *any* of one's *actual* desires is intrinsically good for one. This unrestricted, actualist theory is seldom defended. Perhaps the most common departure from it counts only the satisfaction of *intrinsic* desires, or desires for things for their own sakes, rather than for what they might lead to.[24] When we get what we merely instrumentally want, it is natural to suppose that this is, at best, of mere instrumental value.

Philosophers have considered many other restrictions, such as restrictions to *self-regarding desires* (or desires about oneself),[25] *global desires* (or desires about one's whole life),[26] and *second-order desires* (or desires about one's desires).[27] Some of these will come up when we discuss objections to the desire approach, to which we now turn.

If what's good for us is what we want, then whatever we want is good for us. But surely we sometimes want things that turn out to be no good for us. The most common kind of case involves ignorance. For example, I might have a desire to eat some food, not knowing that it will cause a severe allergic reaction in me, or I might want to see some band perform in concert, not knowing that they will perform terribly. The desire theory seems to imply, mistakenly, that satisfying these ill-informed desires is in my interest.

The lesson that many philosophers draw from such cases is that well-being is connected not to our actual desires but to our *idealized* desires.[28] These are the desires we would have if we knew all the relevant facts, were appreciating them vividly, were making no mistakes in reasoning, and the like. The idealized-desire theory of well-being can claim that it is no benefit to me to eat the allergenic food or attend the bad concert because I would not have wanted these things if I knew the relevant facts and were appreciating them vividly.

Some may hold out hope that the move to idealized desires can solve other problems as well. Perhaps it can provide a solution not only to cases of desires based on mistaken beliefs and the like, but to other sorts of putatively defective desire. Recall the earlier cases of the people who desire to count blades of grass, to break crockery while drunk, or to abuse children. Some desire theorists might be tempted to claim that satisfying these desires is of no benefit because no one who knew all the relevant facts, was appreciating them vividly, was making no mistakes in reasoning, etc. would desire such things. But one has to be careful. This move runs the risk of turning the desire theory into an objective theory of well-being in subjectivist clothing. It is not open to idealized-desire theorists to claim that part of what it is to be idealized is to desire *the right things*, that is, the things it is good to get no matter your desires. That is closet objectivism. The conditions of idealization must be stated in value-neutral terms, and without reference to things that were identified via a belief in their objective welfare value.

Returning to the original objection of putatively defective desires, it is actually not obvious that it succeeds in the first place.[29] Consider the case of the allergenic food. I desire to eat it, not knowing that it will make me sick. The objection claims that the

actualist desire theory is committed to saying that it is nonetheless in my interest to satisfy this desire. But consider two things we might have in mind when we say that it is in my interest to satisfy some desire. We might mean that it is in my interest *all things considered* – that is, taking all the effects of satisfying the desire into account. Or we might mean merely that it is good in itself – intrinsically good – to satisfy the desire. The objection assumes, plausibly, that it is not in my interest *all things considered* to satisfy my desire to eat the food. But the actualist desire theory can accommodate this. For if I satisfy my desire to eat the food, this will cause many of my other actual desires – desires not to be in pain, desires to play golf, etc. – to be frustrated. All that the actualist desire theory is committed to is the claim that it is good *in itself* for me to satisfy my desire to eat the food. But this claim is not implausible. Intuitively, ignoring the effects, it *is* good for me to get to eat this food I very much want to eat.

Thus, moving to an idealized theory may be less well-motivated than it originally appears. It also brings with it new problems. One family of problems concerns the concept and process of idealization.[30] Another problem is that it is possible for what I would want under ideal conditions to be totally uninteresting, or even repugnant, to me as I actually am. But idealized theories of well-being are supposed to tell us what's good for us as we actually are.[31]

The second objection that we will consider has been called the "scope problem" for desire theories.[32] The following example by Derek Parfit illustrates the problem:

> Suppose that I meet a stranger who has what is believed to be a fatal disease. My sympathy is aroused, and I strongly want this stranger to be cured. We never meet again. Later, unknown to me, this stranger is cured. On the Unrestricted Desire-Fulfilment Theory, this event is good for me, and makes my life go better. This is not plausible. We should reject this theory.[33]

James Griffin offers a diagnosis:

> The breadth of the [desire] account, which is its attraction, is also its greatest flaw . . . It allows my utility to be determined by things that . . . do not affect my life in any way at all. The trouble is that one's desires spread themselves so widely over the world that their objects extend far outside the bound of what, with any plausibility, one could take as touching one's well-being.[34]

A common response to this problem is thus that the desire theory should be restricted to count only desires that are about one's own life, or about oneself.[35] According to another proposal, we should count only those desires that are also among our *aims* or *goals* (thus the *aim achievement theory*).[36] Both of these proposals seem to handle Parfit's case. Parfit's desire that the stranger be cured is not about Parfit or his life. Nor is it an aim of

his, since he takes no steps to try to achieve it. Thus each theory agrees that Parfit's life is made no better when the stranger is cured.

But these restrictions may exclude too much. Consider the common desire that one's team win. I do not mean a team one plays for – desires about such a team might count as desires about one's own life, and may qualify as aims – but a team one roots for from a distance. Such desires are certainly not about oneself, and presumably not about one's own life either. And that one's team win is not typically among one's aims or goals; most of us know we have no power over whether our team wins. Thus theories that exclude non-self-regarding desires and desires that are not aims imply, implausibly, that people receive no benefit when their desire that their team win is satisfied.

An alternative solution to the scope problem takes its cue from the detail that Parfit's stranger is cured *unbeknownst to him*. Perhaps the proper scope of the desire theory excludes desires the satisfaction of which we are unaware.[37] This theory gets the right result both in Parfit's case and concerning the desire that one's team win. But it is not clear that it gets to the heart of the initial worry. Here is how T. M. Scanlon presents the initial worry:

> Someone might have a desire about the chemical composition of some star, about whether blue was Napoleon's favorite color, or about whether Julius Caesar was an honest man. But it would be odd to suggest that the well-being of a person who has such desires is affected by these facts themselves.[38]

Scanlon thinks that satisfying such desires is of no benefit even when one is aware that the desires are satisfied. I leave it to the reader to decide whether this is right. Readers may also wish to reconsider the original objection. Some desire theorists maintain that the best reply is to "bite the bullet" in the first place, and maintain that Parfit's life *is* made better when the stranger is cured, even if only a little bit.[39]

The third and final problem for desire theories that we will consider is the problem of changing desires. Our desires change over time. When the desires concern what's going on at the time of the desire, this may be no problem. Each night of the week, I want something different for dinner. In this case of changing desires, the desire theory implies that what's good for me is to get the different meal I want each evening. But some of our changing desires concern what goes on at a single time. Suppose I want, for years, to go skydiving on my fortieth birthday. But as the day approaches, my interests change, and I become strongly averse to doing this.

Probably the most common reaction to this case will be that it is in my interest to satisfy my *present* desire *not* to go skydiving on my fortieth birthday at the expense of frustrating my past desires to go skydiving. (This is assuming that I will not later regret not having gone skydiving – that I will not have persistent desires in the future to have done it.) And this reaction seems right no matter how long held and strong the past

desires to go skydiving were. This suggests that, to determine what benefits a person, we ignore her past desires.[40]

However, other cases might suggest that we *should* take into account past desires. We tend to think that we ought to respect the wishes of the dead – for example concerning whether and where they will be buried. One natural view is that we do this *for their sake* – that is, for their benefit. If that's right, then the desire theory should count at least some past desires. On the other hand, many find it absurd that a person can be benefitted or harmed after he is dead. If that's right, then we must find another explanation for why we should respect the wishes of the dead, assuming that we should.

If past desires can be ignored, this suggests the view that the desire theory count only desires for what goes on at the time of the desire. As R. M. Hare, a proponent of this view, puts it, the theory "admits only now-for-now and then-for-then preferences," to the exclusion of any now-for-then or then-for-now preferences.[41]

But, as before, this might seem to exclude too much. For suppose that I do in fact strongly regret, for years, not having gone skydiving on my fortieth birthday. If so, perhaps it was in my interest to force myself to go skydiving, despite my strong aversion to it at the time, for the sake of satisfying the "then-for-now" desires I would come to have. If that's right, this suggests a surprising asymmetry: the desire theory of well-being should ignore future-directed desires but count present- and past-directed desires. However the problem of changing desires is ultimately resolved, it poses questions that any subjective theory of well-being must grapple with.

Conclusion

The notion of well-being plays some part in answering most, and perhaps even literally all, moral questions. Yet there is no consensus among philosophers concerning which general kind of theory of well-being is correct, or which specific version of any general kind is best. Fortunately, we do not have to know which theory of well-being is correct in order to come to responsible answers to many of the moral questions that involve well-being. For certain kinds of act are harmful and others beneficial on all of the theories of well-being that we have considered. We can thus know that such acts are wrong or right on these bases without having to know precisely what well-being consists in. Still, a full accounting of the act's moral status would require the correct account of well-being.

Thanks to Paul Bowman, Ben Bradley, Ben Eggleston, Eden Lin, Dale Miller, and Jason Raibley.

Notes

1. Sidgwick, *The Methods of Ethics*, p. 420.

2. See Plato, *Euthyphro*, 10a.

3. Nowell-Smith, *Ethics*, pp. 111–113.

4. Railton, "Facts and Values," p. 9.

5. Rawls, *A Theory of Justice*, p. 432.

6. G. E. Moore, *Ethics*, pp. 237–238.

7. Nagel, "Death," p. 77.

8. On hybrid theories, see Parfit, *Reasons and Persons*, pp. 501–502; Kagan, "Well-Being as Enjoying the Good"; and Heathwood, "Welfare," pp. 652–653.

9. Parfit, *Reasons and Persons*, p. 493.

10. See, e.g., Raibley, "Well-Being and the Priority of Values"; and Tiberius and Plakias, "Well-Being," § 3.

11. Mādhava Āchārya, *Sarva-Darśana-Samgraha*, pp. 2–11; and Diogenes Laërtius, *Lives and Opinions*, pp. 81–96.

12. See Epicurus, *Extant Remains*; Bentham, *IPML*, chapter 1 (pp. 11–16); and J. S. Mill, *Utilitarianism*, chapter 2 (*Collected Works*, vol. X, pp. 209–226). See also Aristotle, *Nicomachean Ethics*, book X, chapter 2 (pp. 184–185); and Diogenes Laërtius, *Lives and Opinions*, pp. 89–90.

13. See Sidgwick, *The Methods of Ethics*, book I, chapter 4, §§ 1–2.

14. Feldman, *Pleasure and the Good Life*, chapter 4.

15. Spencer, *The Principles of Psychology*, § 125 (vol. I, pp. 280–281); Brandt, *A Theory of the Good and the Right*, p. 38; and Heathwood, "The Reduction of Sensory Pleasure to Desire."

16. Heathwood, "Desire Satisfactionism and Hedonism."

17. J. S. Mill, *Utilitarianism*, *Collected Works*, vol. x, p. 211.

18. G. E. Moore, *Principia Ethica*, § 47 (pp. 77–79).

19. Nozick, *Anarchy, State, and Utopia*, pp. 44–45.

20. See Railton, "Naturalism and Prescriptivity," p. 170; Crisp, "Hedonism Reconsidered," § 5 (pp. 635–642); and Feldman, "What We Learn from the Experience Machine" for discussions of and/or replies to the experience-machine objection.

21. Feldman, *What Is This Thing Called Happiness?* criticizes life-satisfaction theories of happiness and defends a hedonistic theory, as well as criticizes and defends, respectively, the corresponding versions of eudaimonism. Sumner, *Welfare, Happiness, and Ethics*, chapter 6 defends a life-satisfaction account of happiness and a corresponding eudaimonistic theory of welfare. See also Haybron, *The Pursuit of Unhappiness*.

22. Sidgwick, *The Methods of Ethics*, book I, chapter 9, § 3 (pp. 109–113); Hobbes, *Leviathan*, p. 30; and Spinoza, *Ethics*, part III, prop. 9 (pp. 499–500).

23. See Sumner, *Welfare, Happiness, and Ethics*, p. 113; Shaw, *Contemporary Ethics*, p. 53; and Haybron, *The Pursuit of Unhappiness*, p. 3.

24. Sidgwick, *The Methods of Ethics*, p. 109; and von Wright, *The Varieties of Goodness*, pp. 103–104.

25. Sidgwick, *The Methods of Ethics*, p. 112; and Overvold, "Self-Interest and Getting What You Want."

26. Sidgwick, *The Methods of Ethics*, pp. 111–112; Parfit, *Reasons and Persons*, pp. 497–498; and Carson, *Value and the Good Life*, pp. 73–74.

27. Railton, "Facts and Values," p. 16; and Kraut, "Desire and the Human Good," p. 40.

28. Sidgwick, *The Methods of Ethics*, pp. 110–111; and Brandt, *A Theory of the Good and the Right*, p. 247.

29. Heathwood, "The Problem of Defective Desires," pp. 491–493.

30. Sobel, "Full-Information Accounts of Well-Being"; and Rosati, "Persons, Perspectives, and Full Information Accounts of the Good."

31. Griffin, *Well-Being*, p. 11; see Railton, "Facts and Values," p. 16 for a possible solution.

32. Sumner, *Welfare, Happiness, and Ethics*, p. 135.

33. Parfit, *Reasons and Persons*, p. 494.

34. Griffin, *Well-Being*, pp. 16–17.

35. Overvold, "Self-Interest and Getting What You Want"; and Parfit, *Reasons and Persons*, p. 494.

36. Scanlon, *What We Owe to Each Other*, pp. 119–121.

37. Heathwood, "Desire Satisfactionism and Hedonism," pp. 547–551; and Sumner, *Welfare, Happiness, and Ethics*, pp. 127–128.

38. Scanlon, *What We Owe to Each Other*, p. 114.

39. Lukas, "Desire Satisfactionism and the Problem of Irrelevant Desires."

40. Brandt, *A Theory of the Good and the Right*, pp. 247–253.

41. Hare, *Moral Thinking*, pp. 101–103.

11 Objective theories of well-being

Ben Bradley

According to a common formulation of utilitarianism, an act is morally permissible if and only if it maximizes total well-being. It is important for the utilitarian to investigate the theory of well-being, because many objections to utilitarianism are really objections to a particular theory of well-being rather than to the more general notion that we ought to maximize well-being. For example, the famous "doctrine of swine" objection to utilitarianism is not an objection to the claim that we ought to make people as well-off as we can; it is an objection to the notion that to be well-off is merely to be pleased.[1] There are many theories of well-being, but they are often divided into two sorts: objective theories and subjective theories. This chapter focuses on objective theories. I begin by attempting to explain what makes a theory objective; I then discuss some particular sorts of objective theories.

What is an objective theory of well-being?

Taxonomies of philosophical theories are not inherently interesting. But sometimes it can be useful to see that a bunch of theories have something important in common, so that when we see that a theory has that feature, we will know that the theory is likely to be vulnerable to a particular kind of objection. In this case, put very roughly, subjectivists about well-being argue that all objective theories face a worry about alienation: objective theories tell us that certain things are good for us whether we care about those things or not. But how could something be good for me if I did not care about it? Objectivists, on the other hand, point out that subjective theories entail that we cannot be wrong in what we care about, because there is no objective standard by which to judge our cares; things are good for us merely because we care about them, no matter how worthless, trivial, or immoral those things might be. But surely, says the objectivist, we can be mistaken about what is good for us.

But what is it that makes a theory subjective or objective? This turns out to be trickier than one might think. Here is how Dan Haybron characterizes subjectivism: "Subjectivism about well-being . . . tells us that what ultimately benefits a person is *determined* by subjective psychological states like desires or pleasures."[2]

These remarks might suggest the following distinction:

> *Subjectivism about well-being (version 1):* An individual's well-being is determined entirely by that individual's own subjective psychological states.
> *Objectivism about well-being (version 1):* An individual's well-being depends at least in part on something other than that individual's own subjective psychological

states.

But there are different ways well-being might depend on, or be determined by, subjective psychological states. One way would be for subjective psychological states to themselves *be* the constituents of well-being. Haybron mentions pleasure as an example of the sort of psychological state that a subjectivist thinks determines well-being. Someone who thinks that pleasures determine one's level of well-being thinks that pleasures are themselves the fundamental constituents of well-being. On the other hand, subjective psychological states might determine well-being by *picking out* the constituents of well-being. If desire is the relevant psychological state, then what is good for someone is the object of the desire, not the desire itself. These are importantly different ways a psychological state can determine well-being, and it is the second way that subjectivists typically have in mind.

Before making a second pass at the distinction, let us distinguish between different sorts of subjective psychological states. On the one hand, there are *feelings*, such as heat, coldness, pressure, and (perhaps) pleasure and pain. On the other hand, there are *attitudes*, such as belief, desire, fear, hope, and the like. Attitudes are *about* something. Sometimes they are about propositions, such as when Jeff believes *that his pants are on fire*; sometimes, perhaps, they are about objects, such as when Buffy desires *a new bowling ball*. (Perhaps when the object of a desire seems to be a physical object such as a bowling ball, we should really think that it is a proposition, such as *that I have a bowling ball*. Perhaps it is easier to see how Buffy's having a bowling ball could be a constituent of Buffy's well-being than it would be to see how the ball itself could be a constituent of her well-being.) In this way attitudes are unlike feelings. The feeling of heat may be caused by a fire, but it is not *about* the fire.

It is natural to think that both feelings and attitudes are among the psychological states that a subjectivist thinks are relevant to well-being, and this is suggested by the quotation from Haybron. In fact, however, it is more typical for the subjectivist to say that only attitudes, and not feelings, are directly relevant to well-being. So, for example, consider how the subjective/objective distinction is drawn by L. W. Sumner and Richard Arneson:[3]

> A subjective theory will map the polarity of welfare onto the polarity of attitudes, so that being well off will depend (in some way or other) on having a favourable attitude toward one's life (or some of its ingredients) . . . On an objective theory . . . something can be . . . good for me though I do not regard it favourably, and my life can be going well despite my failing to have any positive attitude toward it.[4]
>
> I would prefer to let the contrast between objective and subjective mark the contrast between (1) views which hold that claims about what is good can be correct or incorrect and that the correctness of a claim about a person's good is determined independently of that person's volition, attitudes, and opinions, and (2) views which

deny this.[5]

Let us then understand subjectivism and objectivism in the following way:

> *Subjectivism about well-being (version 2):* All the things that are good for an individual are good for her in virtue of her attitudes about them (e.g., in virtue of the fact that she desires them for their own sakes).
> *Objectivism about well-being (version 2):* Some of the things that are good for an individual are good for her independently of her attitudes about them.

This will do for a start. But it may be easier to tell where the distinction should be drawn after we have seen examples of objective theories and how the subjectivist objects to them. So at the end of this chapter I will return to the classificatory question. I turn now to some specific sorts of objective theories.

Hedonism

Hedonism is the view that the only components of well-being are pleasure and pain.[6] Pleasure is the sole positive welfare component, and pain is the sole negative welfare component. How well someone's life goes for her is determined by subtracting the total amount of pain she receives in her life from the total amount of pleasure she receives. At first blush, hedonism seems to be a form of objectivism, because pleasure makes one's life go better, and pain makes it go worse, no matter what one's attitude toward the pleasure or pain might be.

On one way of thinking about pleasure, however, hedonism turns out to be a version of subjectivism. On this view, what makes a feeling a feeling of pleasure is that the person having the feeling desires that it continue.[7] This makes hedonism a sort of subjectivism, because whether a feeling is a pleasure, and hence whether that feeling contributes to well-being, depends on the attitude of the person having the feeling. This is an attractive view; it is certainly true that, in general, we want our pleasures to continue. But do we always want this? And couldn't there be feelings that we want to continue even though they are not pleasures? These are difficult questions. For the purposes of this chapter I will assume that pleasure is a distinct sort of feeling, not reducible to desire; thus I will take hedonism to be a sort of objectivism.[8] I merely wish to flag the interesting point that, if hedonism is true, whether subjectivism or objectivism about well-being is true might depend on how we answer these difficult questions about the nature of pleasure.

Hedonism has a good deal of initial plausibility and explanatory power. First, it is very plausible that at least some pleasures increase our well-being, and that at least some pains decrease it. This perhaps puts some pressure on the anti-hedonist to explain how it could

be that some pleasures and pains do not affect our well-being in this way. Second, concerning many of the things that we think of as being good or bad in some way, a plausible story can be told that links those things to pleasure or pain. This perhaps puts pressure on the anti-hedonist to show that there are some things about which no such story can be told. So anti-hedonists have primarily given two sorts of objections to hedonism: that not all pleasures are good for us, and that some things other than pleasures are good for us.

It is often claimed that *false pleasures* are not good for us. Robert Nozick's well-known example of the "experience machine" is sometimes employed to show this.[9] Imagine you could be hooked up to a machine that would give you any experiences you wanted to have. While in the machine you would think that everything that was happening was real (as in many science fiction movies). You would think you were dating famous models, playing center field for the Dodgers, climbing Mount Everest, or whatever would give you the most pleasure. Of course, you would not be doing any of those things. Many people say that they would not choose to be hooked up to such a machine. Perhaps they think that the pleasures they would get while in the machine would not be valuable. We can also think of less fanciful deceptions; if someone believes her spouse loves her and takes pleasure in this belief, while in fact, unbeknownst to her, her spouse only married her for the money and is cheating on her at every opportunity, we might think that the pleasure she takes in her marriage is not valuable to her.

It is also often claimed that *immoral pleasures* are not good for us. For example, the serial killer who derives great pleasure from torturing and murdering his innocent victims, and never gets caught or feels guilty about his actions, is not thereby well-off.

Hedonists might not be bothered by such objections. Regarding immoral pleasures, they might argue that the serial killer's pleasures *do* benefit him – and that in fact this partly explains why the situation is so appalling, since he so obviously does not deserve to be well-off.[10] Regarding false pleasures, they might stick to their guns and say that ignorance can be bliss, or that what you don't know can't hurt you. Or they might say that, although people would prefer not to be plugged into an experience machine, this is not because they think they would be worse off if they did; rather, it is because they prefer to sacrifice some of their own well-being for the sake of other things, such as genuine relationships, commitments, or knowledge.[11]

But there are other options for the hedonist. Mill famously claimed that pleasures come in different qualities, in addition to different quantities.[12] The pleasures of doing philosophy, enjoying artwork, or acting morally are higher-quality pleasures than those of, for example, eating food. Perhaps Mill would have thought of immoral and false pleasures as low-quality pleasures. More recently, several philosophers have endorsed impure or hybrid forms of hedonism in an effort to avoid these objections; I discuss these views below.

There is another, stronger objection to hedonism. It seems that some things other than pleasures are good for us. To see what sorts of things those might be, it is helpful to think about the experience machine again. What would a life on the experience machine be lacking? There are many candidates, including knowledge, achievement, virtue, and friendship. These things all seem to be good for us. However, the hedonist will argue that such things are good for us only if, and to the extent that, they improve the hedonic status of our lives. Imagine someone who has immense knowledge about the universe, but gets no enjoyment at all from this knowledge; or someone who has many friends, but does not enjoy any of these friendships even a little bit. Are these people well-off?

These arguments and replies about the good life have been given at least since Plato's *Philebus* and probably longer than that. We will not resolve them here, so let us now turn to other objective theories, and see what sort of objective view we might hold if we are convinced by the anti-hedonist arguments.

Objective list theories

Why think that only one thing can be good for us? Why not two, or ten? According to what is sometimes called the objective list theory, or pluralism, there are several things that are good or bad for us; when we arrive at a list of these things, we have reached rock bottom in our investigation into the fundamental elements of well-being.[13] No further explanation is possible.[14]

It is perhaps misleading to refer to "the" objective list theory. For there are many possible lists. The "theory" has no content until we identify the items on the list. Pleasure and pain are obvious candidates for the list. Other prominent candidates include achievement and failure, knowledge and false belief, virtue and vice, and friendship and loneliness. The objective list theorist might also wish to add what we might call "second-order" goods and evils to the list. For example, we might think that it is better to have a *variety* of goods in one's life than it is to have a similar quantity of very similar goods. Or we might think that it is better to have a life that *improves* over time than to have one that deteriorates, even if the sum of goods is the same: better to have the bad things in one's life closer to the beginning, and the good things toward the end, than vice versa.[15] Since it is an objective list, all these things would be good or bad for someone no matter what that person thinks about it.

The objective list theory has some obvious strengths. It is immune to the objection that it leaves something out of the good life. Thus it seems to be invulnerable to, e.g., arguments based on experience machines and such. If you think something has been left out of the theory, then just add that thing to the list. As a result, many find the objective list theory more intuitively plausible than hedonism or other monistic theories. There are certain sorts of cases where it seems particularly important to employ an objective list

theory. For example, consider a society in which women are systematically oppressed: they cannot go outside without being accompanied by a man, cannot go to school, cannot have a job outside the home, etc. This seems like a deprived existence. But sometimes people can adapt to such situations. They sometimes find a way to be happy and satisfied even while being oppressed; they might even come to prefer such a lifestyle to one in which they would have more opportunities. Still, we want to say that not everything is all right. Even though they would not agree, their lives would be better if they had more freedom to acquire knowledge and achieve things. Yet from the perspective of the content-but-oppressed person, everything is fine and nothing is lacking. This shows, we might think, that our perspectives and our attitudes can be wrong, and that one can be inappropriately pleased or satisfied with how one's life is going.[16]

However, there are certain questions that it seems difficult or impossible for the objective list theory to answer. The most obvious question is why these things, rather than some other things, are on the list. This question cannot, in principle, be answered by the objective list theory. There is just a list; that is the end of the story.

But is this a problem for the list theory? According to Roger Crisp there is a difference between "enumerative" and "explanatory" theories of well-being.[17] The list theory would count as an enumerative theory, because it tells us *which* things are good for us. It would not count as an explanatory theory, because it does not tell us *why* those things are on the list. We might take it as a defect of the list theory that it does not count as explanatory. But Crisp's distinction is ultimately not very helpful in identifying a special problem for objective list theories relative to other theories. Once we arrive at a fundamentally good thing, the only explanation of what it is that makes that thing good for us is *that it is the sort of thing that it is*. This is so whether it is the only fundamentally good thing or one of several. Thus, as Crisp says, "the hedonist . . . will say that what makes accomplishment, enjoyable experiences, or whatever good for people is *their being enjoyable*."[18]

The objective list theorist will point out that explanation must stop somewhere; where explanation stops, the objective list theorist finds more than one thing. We should not look for deeper unity where there is none to be found. She might also point out that hedonism is not in an obviously better position on this score: why is only this one thing, pleasure, on the good list? The hedonist cannot answer this question either. Perhaps the objection to the objective list theory relies on the natural thought that, for certain sorts of things at least, the question "How many of these things are there?" has some answers that seem less arbitrary than others. Zero, one, and infinitely many are non-arbitrary-seeming answers; 3, 8, and 745,982 are arbitrary-seeming answers.[19] But this depends on the sort of thing we are talking about. It is not arbitrary to say there are two kinds of elephant or 1,467,358,448 kinds of insect. That is just how many there are (let us imagine). We figured it out by counting them. Maybe we can just count up the intrinsic

goods, and find out that there are five kinds. But is this a sufficient reply on behalf of the list theory? The fact that there are two kinds of elephant is not a brute, unexplainable fact. Evolutionary biologists can tell us a story about why there are two kinds of elephant. But if there are exactly five sorts of things that are good for a person, this is, on the objective list story, a brute, unexplainable fact about the universe. This may be hard to believe.[20]

But we should not rule out the theory on the basis of this general consideration before we look more closely at the particular candidates for the list. Begin with knowledge: are we better off just for knowing things, and are we worse off just for having false beliefs about things? If I think there are a million blades of grass in my yard, but in fact there are a million and one, am I worse off? Suppose I counted the blades and discovered that there were a million and one, and so I have a true, justified belief about the number of blades of grass in my yard; could my neighbor then make me worse off by, unbeknownst to me, removing a blade of grass from my yard, thereby rendering one of my beliefs false (even if justified)? And would his putting it back restore my well-being? This will strike some as unlikely. But perhaps, just as the hedonist might try to distinguish worthy pleasures from worthless ones, the objective list theorist could try to distinguish worthy knowledge from worthless knowledge.[21] How many blades of grass are in my yard is not worth knowing, but the fundamental physical laws of the universe are worth knowing. Of course, the hedonist has an explanation for this: knowing the fundamental physical laws is more likely to promote happiness than knowing how many blades of grass are in my yard. But is this the only thing that makes it worthwhile to know such things? We often behave as if we do not think so. People pursue knowledge even when they do not believe it will result in any pleasure for them or others. We donate money to universities and build expensive machines to explore outer space. Maybe we do these things for the sake of pleasure, but it is far from clear that this is so, since there are other things we could do that seem likely to promote pleasure or prevent pain more effectively. Of course, some would argue that if this is so, we should be doing those other things. Rather than use resources to fund universities or explore space, we should be preventing suffering by donating to Oxfam and such.

Achievement is another candidate for the list. To achieve something is, roughly, to put forth some efforts toward a goal, and for those efforts to be successful. Once again there seem to be things worth achieving and things not worth achieving; finding a cure for a deadly disease is worth achieving, while reaching level 50 on Angry Birds is not.[22] So perhaps it is implausible to say that achievement per se is intrinsically good for us. The objective list theorist might say that it is better, for one who is trying to reach level 50 on Angry Birds, to be successful than to fail.[23] Or it could be argued that the value of an achievement depends on the value of what is achieved. But is it good to achieve something, valuable or not, if one would not be at all pained by failing, and would get no pleasure from succeeding?

Virtue is another plausible candidate. Certainly we want those we care about to be (at least to some extent) courageous, honest, and kind people. But this might just be because we think being virtuous is likely to make one better off instrumentally, by making other people like one more, etc. Or we might care that someone is virtuous without thinking that virtue is intrinsically good *for him*; we might think the universe is better for having virtuous people in it.

One way to try to figure out whether something is an element of well-being is to think about reward and punishment.[24] When someone has done something bad, we might try to punish that person by doing something that would negatively impact his well-being. Those who think punishment ought to make someone worse off might find it appropriate to punish someone by inflicting some pain, or preventing some pleasure, or preventing the person from getting what he wants. But it seems totally inappropriate to punish someone by making him more cowardly, dishonest, or miserly. Perhaps it is similarly strange to think of rewarding someone for a good deed by making her more beneficent. We might take this to be a reason to doubt that virtue and vice are components of well-being.

What about friendships and, more generally, loving relationships? These are perhaps the most important things one would miss in the experience machine; one would believe one had loving relationships with one's family and friends, but in fact one would have no such relationships. It is hard to see such an existence as being wonderful. On the other hand, friendships that bring no enjoyment seem rather pointless. We sometimes have to decide whether to put effort into maintaining a relationship with someone. Sometimes these decisions are not made on the basis of self-interest; we might maintain a relationship out of a sense of moral duty. But when this is not the case, it seems reasonable to decide not to maintain a relationship that is bringing only annoyance or pain. (This might point us toward a hybrid view of well-being; see the "Hybrid Views" section below.)

Suppose we find a list of goods we are happy with. Our work will still not be done. For another question remains: how are the items on the list to be weighed? We must have some way to determine how well someone's life goes, given that it has some combination of those goods.

A flatfooted thought is just to add up the values of the items on the list. For example, if Joe gets 10 units of pleasure, and 10 units of knowledge, and 10 units of virtue, his life would have a value of 30 for him. But how are the "units" to be determined across types of value? The choice of unit will be doing all the work here. The question of how to weigh the items on the list merely gets reformulated as the question of how to determine what counts as one unit of each sort of thing.

One thought would be that there is a *lexical ordering*.[25] For example, suppose that there are only two goods: pleasure and knowledge. We might think that pleasure always

outweighs knowledge. When comparing two possible lives, we first look at which one contains more pleasure, and if one contains more, then it is better; if they are equal in pleasure, we then look at which one contains more knowledge. But this is very implausible; if knowledge really impacts well-being, how could it be that a tiny bit of pleasure could outweigh an enormous amount of knowledge? Since the amounts of pleasure in any two lives would almost never be identical, knowledge would effectively have little or no impact on well-being.

If there is no lexical ordering, what determines the relative values of the good things? This question might be impossible to answer for the objective list theorist. But once again, the hedonist has a similar problem: how can the positive values of pleasures and the negative values of pains be compared? Pleasure and pain are distinct feelings, and it is not clear how to compare one combination of pleasure and pain with another. This seems to be a problem for all objective views, but perhaps not a problem for subjective views. If what is good for an individual is getting what she wants, and what is bad is not getting what she wants, the goodness of getting what one wants and the badness of not getting what one wants may both be determined by the same thing: the degree of desire.

While it is initially plausible to say that such things as knowledge, achievement, virtue, and friendship are intrinsically good for us, we can see that it is also plausible to say that they are merely very important instrumental goods. Rather than just appeal to intuitive judgments about the values of these things, we might try to find a more general reason to think that these things must be on the list: an organizing principle for the list that tells us what goes on the list and why. Perfectionism promises to provide such an account. So we turn now to perfectionism.

Perfectionism

A promising thought about well-being is that what is good for us is determined by *what kind of thing we are*. But of course we are many kinds of thing: we are physical objects, living things, medium-sized things, intelligent and sentient things, walking things, etc. What is good for us cannot be determined by every sort of thing we are. Some of these kinds are special. They are the kind of thing we are in a more *fundamental* way. Right now I am a sitting thing, but this has little to do with making me what I am. Shortly I will be a standing thing, but I will not have undergone any fundamental change. Perfectionism requires us to begin with a metaphysical question: what is the *nature* of a human being? When we have answered that question, then perhaps we can derive the answer to what is good for us from it: to be well-off is to perfect one's nature. Perhaps perfecting one's nature will involve such things as acquiring knowledge, achieving things, having friends, and enjoying oneself. Perfectionism promises to tell us what is going wrong with people who are contentedly oppressed: in a society in which some people are treated as less than fully human, those people fail to perfect their natures as humans, which is a tragedy for them.

It is no easy task to say what our nature is, and of course many will be skeptical that there is such a thing as human nature at all. As a first pass, we might say that our nature consists of the properties that are essential to us. This would rule out, e.g., being a sitting thing as being part of our nature. But it might also rule in other properties, such as being a physical object, that seem irrelevant to what is good for us. What would it be to perfect one's nature as a physical object? Some properties are essential to us, but do not make us *what we are*, at least in the sense we care about when thinking about what is good for us.

Thomas Hurka has defended a perfectionist view according to which what is valuable is the development of those properties that are essential to human beings "qua living things," or properties that are both essential to humans and distinctive of living things.[26] According to Hurka, this version of perfectionism entails that what is good for us is to develop our physical nature (to be healthy and physically fit), our theoretical rationality, and our practical rationality.[27] This seems to do a nice job of capturing at least some of the items on the pluralist's list of goods; for example, developing one's theoretical rationality involves acquiring knowledge; developing one's practical rationality involves acquiring (at least some) virtues.

We might wonder, though, whether it can capture all the goods that belong on the list. For example, pleasure seems like it ought to be on the list, but one can be healthy and rational without enjoying oneself. Perhaps the best that can be said for perfectionism is that it adds some unity to the list, which is nice, but does not completely unify the list.

A deeper problem for Hurka's view is that it seems to rely on an implausible view about what is essential to humanity. There are members of *homo sapiens* that do not have the properties identified by Hurka as essential to humanity: those with serious brain damage, and very young infants, for example, do not have the sort of rationality that is alleged by Hurka to be part of the human essence.

Furthermore, it is hard to see how to avoid ruling in some things that we do not want to rule in. Human beings can be cruel to one another and to other beings. Suppose cruelty turns out to be part of human nature; then it would turn out to be good for us to be cruel. This seems hard to accept.[28]

Hybrid views

Recall that one problem for hedonism was the problem of worthless pleasures: some things seem to be the wrong sort of thing to enjoy. Recall also that a problem for perfectionism is that it leaves pleasure out of the picture, and that a problem for, e.g., the view that friendship or knowledge is intrinsically valuable is that friendship or knowledge that brings no enjoyment seems pointless. We might try to solve all of these problems with a hybrid view. According to hybrid views, what is good for someone is to take

pleasure in something that is worthy of having pleasure taken in it, or to want something good and get it. Several philosophers have endorsed or flirted with this sort of view, including Derek Parfit, Susan Wolf, Fred Feldman, Robert Adams, Stephen Darwall, Shelly Kagan, and Richard Kraut.[29] The hybrid theory is "subjective" enough to avoid at least some worries about alienation such as those raised at the start of the first section of this chapter, because getting some allegedly good thing will not be deemed by the theory to benefit one unless one wants or enjoys that thing. But it is still an objective theory, because what things are worth enjoying or desiring is an objective matter. Furthermore, the hybrid view can account for the intuition that virtue is a component of well-being. As Shelly Kagan points out, on one way of thinking about virtue, virtue consists of loving the good; and one way to love something is to enjoy or desire it; so when one enjoys what is good, one is thereby virtuous.[30]

But which are the sorts of things that are worthy of being enjoyed or desired? Here we face familiar problems. Perhaps there is just a list of things that are worthy of being enjoyed. Or perhaps the things that are worthy of being enjoyed are instances of perfecting someone's nature.

Hybrid views also face another sort of problem.[31] Suppose we think that the better an object of enjoyment is, the better it is to enjoy it.[32] Some things seem appropriate objects of enjoyment, but have *indeterminate* values. For example, it is appropriate for me to enjoy the fact that my son is enjoying himself to some extent. But my son enjoying himself to some extent has indeterminate value. So how good is it for me that I enjoy my son's enjoyment? It seems that the value of my enjoyment must be indeterminate for me. This would present complications for the consequentialist who thinks we ought to maximize well-being, for it will sometimes be indeterminate which option maximizes well-being, and therefore indeterminate whether some action is permissible.

Alienation

As mentioned in the first section of this chapter, the subjectivist has an argument against all the versions of objectivism stated so far. She says: How can pleasure, knowledge, virtue, perfecting one's nature, or anything else be good for me if it is not something I care about? The perfectionist says I will be better off if I perfect my human nature. But I do not care about my human nature, and why should I? How can I be wrong in failing to care about it?[33]

If this is supposed to be an *argument* against objective theories, it is fair to ask why it should be convincing. Once we have grasped the distinction between objective and subjective accounts of well-being, the alienation argument amounts to little more than pointing to the fact that objective theories are not subjective theories, and saying: "See? Objectivism is not true, because it is not subjectivism." But subjectivism does have

intuitive appeal. When we think of someone living a life doing things she is not interested in and does not care about, it is hard to think of her life as being good for her. The good life is supposed to seem attractive.

The strength of this argument might depend on the particular sort of objective theory it is leveled against. For example, if someone just does not care about knowing things unless this helps her in some other way, it might seem right to say that knowing things is not good for her. On the other hand, if someone does not want to experience any pleasure or does not care about avoiding pain, we might well think something is wrong with her desires. "Why wouldn't she want to avoid pain?" we might ask. This suggests that the alienation argument seems strong when leveled against a questionable candidate for the list, but loses force when employed against a more plausible candidate.

Reconsidering the distinction

I now return to the distinction between objective and subjective theories. Recall that we distinguished subjectivism and objectivism in the following way:

> *Subjectivism about well-being (version 2):* All the things that are good for an individual are good for her in virtue of her attitudes about them (e.g., in virtue of the fact that she desires them for their own sakes).
> *Objectivism about well-being (version 2):* Some of the things that are good for an individual are good for her independently of her attitudes about them.

There are reasons to think that this way of distinguishing the theories cannot be right. So I would like to suggest a tempting revision, even though it may also be unsatisfactory.

Consider the following view about well-being. What is good for someone is wanting something and getting it; what is bad for someone is wanting something and not getting it. This is a version of desire-satisfactionism about well-being. It sounds like a subjective theory of well-being, and it is not subject to any worries about alienation. But given version 2, it counts as an objective theory. This is because it attributes value to the combination of my wanting something and getting it, and that combination is good for me no matter what my attitude is about that combination.

The reason this is interesting is that there are two distinct versions of desire satisfactionism, the "object view" (according to which it is *the thing desired* that is good) and the "combo view" (according to which the *combination* of desiring the thing and getting the thing is good), that give us exactly the same results concerning which lives contain more well-being than which – but given version 2 of our distinction between theories, one of those is a subjective view and the other is objective. The difference between the views is to be found not in how much value there is in a given life, but in *where that value is located*: in the objects of the person's desires, or in combinations of

the person's desires and their objects. And we might well wonder, given that this is the case, why we should worry about the distinction between objective and subjective theories. The mere fact that one theory is subjective and another theory is objective does not necessarily lead to any distinction in how they evaluate lives. In fact, it might be that every subjective theory has an objective counterpart that yields the same results.

We might think that this shows that version 2 draws the distinction between subjective and objective views improperly. We might be tempted to redraw the distinction to get these desire-satisfactionist views on the same side of the divide. Here is one way we might do that:

> *Subjectivism about well-being (version 3):* All the things that are good for an individual involve that individual's desires in some way.
> *Objectivism about well-being (version 3):* Some of the things that are good for an individual do not in any way involve that individual's desires.

It is important that we restrict the relevant attitudes to desires; otherwise we would risk misclassifying as subjective, e.g., the view that knowledge is intrinsically good, since knowledge involves an individual's attitudes (namely her beliefs). Version 3 gets both the combo and object versions of desire satisfactionism on the subjectivist side of the divide. The combo view counts as subjective because the good things include desires as parts; the object view counts as subjective because the good things are objects of desires. But does version 3 include too many theories as subjective? Consider the following wild view: what is intrinsically good for me is the combination of some item on the objective list and my desire that 2 + 2 = 4. According to this view, what is good for me involves my desires in some way. But it is *the wrong way*, from the standpoint of a subjectivist. This wild view would be subject to all the worries about alienation that motivate the move to subjectivism in the first place. So the challenge in formulating subjectivism is to say something sufficiently specific about how an individual's desires are related to what is good for her, without saying something so specific that, e.g., combo views do not count as subjective.

It might be that philosophers – subjectivists, in particular – are led to believe that there is an important distinction between objective and subjective theories of well-being for reasons having little to do with judgments about what sorts of lives are valuable. They might rather be motivated by metaphysical and epistemological concerns. A likely motivation for being a subjectivist about well-being is a commitment to a naturalistic metaphysical worldview that does not allow for an irreducible, non-natural property of goodness for an individual (this may be in part because of epistemological worries about such properties). If you are suspicious of such properties, but you believe that some things are good for you and some things are bad for you, it is natural to think that all this amounts to is that you want or like some things and do not want or like others. Certain things seem to have the glow of goodness about them, but this just means we want them;

other things seem to have the stench of badness emanating from them, but this just means we are averse to them.

If we had subjectivist leanings, we might find the combo view problematic. It seems to require a property of goodness for an individual that is not itself reducible to wanting or liking, even though it states that wantings are part of what is good for an individual. The object view does not require such a property; one who defends the object view can say that the property of being *good for* an individual just is the property of being *desired by* that individual. Now, the defender of the combo view might try to say something similar: that the property of being good for an individual *just is* the property of being a combination of desiring something and getting that thing.[34] But the glow of goodness and the stench of badness do not seem to emanate from such combinations, as they seem to emanate from things we desire and things to which we are averse. In light of this, the subjectivist may wish to stick to her guns and place the combo view on the objectivist side of the divide after all.

Thanks to Ben Eggleston, Chris Heathwood, Eden Lin, and Dale Miller for helpful comments on a previous draft. Thanks also to Dale Dorsey, Kris McDaniel, and David Sobel for helpful discussion of the arguments discussed herein.

Notes

1. J. S. Mill, *Utilitarianism, Collected Works*, vol. x, p. 210.

2. Haybron, *The Pursuit of Unhappiness*, p. 13; also see Griffin, *Well-Being*, p. 33; and Ferkany, "The Objectivity of Wellbeing," pp. 474–475.

3. Also see Griffin, *Well-Being*, p. 32; and Parfit, *Reasons and Persons*, p. 499.

4. Sumner, *Welfare, Happiness, and Ethics*, p. 38.

5. Arneson, "Human Flourishing versus Desire Satisfaction," p. 115.

6. For recent defenses of hedonism, see Goldstein, "Pleasure and Pain"; Feldman, *Pleasure and the Good Life*; Crisp, "Hedonism Reconsidered" and *Reasons and the*

Good; Mendola, *Goodness and Justice*; and Bradley, *Well-Being and Death*.

7. Sidgwick, *The Methods of Ethics*, p. 42; see Heathwood, "The Reduction of Sensory Pleasure to Desire," for a recent defense of this sort of view.

8. On the view of pleasure espoused by Fred Feldman in *Pleasure and the Good Life*, pleasure is not a feeling at all, but an attitude. But it is the having of the attitude that is good for someone, not the object of the attitude. So his version of hedonism counts as objective given version 2 of the distinction between subjective and objective views.

9. Nozick, *Anarchy, State, and Utopia*, pp. 42–45.

10. Goldstein, "Pleasure and Pain."

11. Kawall, "The Experience Machine and Mental State Theories of Well-Being."

12. J. S. Mill, *Utilitarianism*, ch. 2 (*Collected Works*, vol. x, pp. 209–226).

13. As Eden Lin has pointed out to me, in principle there is no reason one could not be a pluralist and a subjectivist, or a monist and an objectivist (hedonism is an example of the latter). So it is not ideal to identify the objective list theory with pluralism. But the name 'objective list theory' has attached to pluralist views, for better or worse, thanks to Parfit's influence (Parfit, *Reasons and Persons*, p. 493).

14. See W. D. Ross, *The Right and the Good*, chapter 5 (pp. 134–141), for a classic objective list view.

15. C. I. Lewis, *The Ground and Nature of the Right*, p. 68.

16. This sort of argument has been widely discussed; for one example, see Nussbaum, "Adaptive Preferences and Women's Options."

17. Crisp, *Reasons and the Good*, pp. 102–103.

18. Crisp, *Reasons and the Good*, p. 103.

19. See David Lewis's discussion of possible sizes of spacetime (*On the Plurality of Worlds*, p. 103).

20. Thanks to Chris Heathwood for discussion of this thought. In principle there might be no reason the objective list theorist could not give a separate explanation, for each item on the list, of why that item is on the list. But the presence of an item on the list is usually taken to be not subject to further explanation.

21. W. D. Ross, *The Right and the Good*, pp. 145–149.

22. Note to readers of the distant future: long ago, Angry Birds was a video game popular with children (I am told).

23. Keller, "Welfare and the Achievement of Goals."

24. See also Brad Hooker's "Sympathy Test" ("Does Moral Virtue Constitute a Benefit to the Agent?" pp. 149–155).

25. For an example of a lexical view see W. D. Ross, *The Right and the Good*, chapter 6 (pp. 142–154); Ross is concerned with goodness *simpliciter* rather than well-being but the issues are the same.

26. Hurka, *Perfectionism*, p. 16.

27. Hurka, *Perfectionism*, p. 37.

28. See Dorsey, "Three Arguments for Perfectionism," for a recent critical discussion of arguments in favor of perfectionism.

29. Parfit, *Reasons and Persons*, pp. 501–502; Wolf, *Meaning in Life and Why It Matters*; Feldman, *Pleasure and the Good Life*; Adams, *Finite and Infinite Goods*, pp. 93–101; Darwall, "Valuing Activity"; Kagan, "Well-Being as Enjoying the Good"; and Kraut, "Desire and the Human Good."

30. Kagan, "Well-Being as Enjoying the Good," pp. 261–262.

31. For a detailed explanation of this argument, see Lemos, "Indeterminate Value, Basic Value, and Summation."

32. Kagan, "Well-Being as Enjoying the Good," pp. 263–264.

33. Railton, "Facts and Values," p. 9.

34. Thanks to Chris Heathwood for pointing this out to me and to Eden Lin for helpful clarification.

12 Kantian ethics and utilitarianism

Jens Timmermann

Kant's ethics and utilitarianism: historical rivals

Immanuel Kant's ethical theory is often considered the most important modern rival to utilitarianism. Both theories are products of the same era in that their foundations were laid during the final decades of the eighteenth century. Jeremy Bentham completed his manuscript of *An Introduction to the Principles of Morals and Legislation* in 1780. It was published in 1789, shortly after Kant's foundational ethical works, the *Groundwork of the Metaphysics of Morals*, which appeared in print in 1785, and the *Critique of Practical Reason* of 1788. A full statement of Kant's legal and ethical theory, the *Metaphysics of Morals*, was to follow in 1797.

Two facts about the emergence of Kantian and utilitarian ethics are particularly striking. First, while their proposals are (arguably) very different, Bentham and Kant are naturally understood as addressing the same philosophical question: What is the principle of morality, the highest standard of what human beings ought to do? Their candidates are the greatest happiness principle and the categorical imperative, respectively. Note that both moral theorists share an assumption that is by no means uncontroversial: that there is such a supreme principle. Second, even though both theories were developed during the same decade they emerged not only on opposite sides of the English Channel but completely independently of each other. Kant classified all other ethical theories known to him and dismissed them as incompatible with the autonomy of the human will, for him the only basis of an account of moral obligation, but he was oblivious of the existence of Bentham's rival principle. Likewise, in developing his own ethical system, Bentham did not engage with Kant's proposed categorical imperative.

Later generations of utilitarians and Kantians did of course interact, and their mutual criticisms can be very instructive. A particularly prominent early example is John Stuart Mill's discussion of the categorical imperative in *Utilitarianism* (1861). Mill mentions the Kantian rival principle twice: in the general remarks in chapter 1 and in his discussion of justice and utility further down in chapter 5. What is more, Mill's first chapter contains several statements concerning his ethical methodology that bear fruitful comparison with Kant's strategy in the *Groundwork*.

Methodological assumptions: the *summum bonum* and everyday morality

Let us start by examining Mill's and Kant's methodological assumptions. On the first

page of *Utilitarianism*, Mill complains about the "confusion and uncertainty" that exists with regard to the first principle of ethics.[1] He explicitly equates the question concerning the foundation of morality with the search for the *summum bonum*, the highest good, which in some shape or form had dominated moral philosophy since antiquity. According to Mill, the specific conception of the highest good that is presupposed – according to him: the happiness of all, considered impartially – determines the right and wrong of human conduct. Actions are right if they produce the greatest possible amount of happiness.

In the *Groundwork*, Kant similarly operates on the assumption that the subject matter of ethics is defined by the classical problem of the highest good. On the first page of the first section he famously presents his own candidate: a morally good will.[2] But for Kant the good will is not the only intrinsic good, the other being happiness. It is the supreme good and the only unconditional good. The goodness of valuable things other than the good will depends on their being in harmony with good willing. If the presumed value of something is in conflict with the good will it does not have objective value. In the eyes of reason it is worthless. All things considered, something that contradicts morality is not good at all.

Kant proceeds to consider happiness as a rival candidate for the accolade of the highest good. Like classical utilitarians he construes it along experientialist lines as "the entire well-being and contentment with one's condition,"[3] the result of seeing one's desires satisfied. But he immediately dismisses happiness. It is true, Kant concedes, happiness *can* be objectively good. Everyone wants to be happy. It is not valued merely as a means. As we shall soon see in more detail, we even have a duty to make others happy. But happiness is good only when certain conditions are met.[4] To be precise, Kant argues that happiness must be earned: only morally good people are worthy to be happy.[5] The happiness of the wicked has no objective value. Indeed

> a rational impartial spectator can nevermore take any delight in the sight of the uninterrupted prosperity of a being adorned with no feature of a pure and good will, and . . . a good will thus appears to constitute the indispensable condition even of the worthiness to be happy.[6]

If this seems counterintuitive, consider an analogy from the world of sports. Winning is the inherent purpose of engaging in any competitive sport, just as being happy is the inescapable purpose of any human being alive.[7] But Kantians would argue that being desired is insufficient to make achieving either victory or happiness good, i.e., objectively worth having. In the judgment of an honest sportsperson victory is worth having only within the rules of fairness, just as in the judgment of Kant's pure practical reason happiness is deserved only within the rules set by morality. It is bad that someone should win the gold medal as a result of an advantage gained by the use of illegal substances.

Such a victory may seem attractive. Objectively it is still worthless. Kant would argue that the happiness that results from arrogating to oneself and one's desires a status one does not have – e.g., by making an exception for oneself at the expense of others, violating the basic equality of status we all share – is worthless in very much the same way. No doubt, both undeserved victory and underserved happiness can be pleasant, but they lack the approval of impartial reason. As a result they are not good.[8]

In sum, there seems to be common ground between utilitarianism and Kantian moral philosophy with regard to the task at hand. Both Mill and Kant are looking for a supreme principle of ethics, both think that the question can be framed in traditional terms, with reference to a *summum bonum*, and both consider happiness a natural candidate, construed along similar experientialist lines. Yet they offer very different answers as to what they take the highest good to be. Mill embraces happiness whereas Kant rejects it in favor of morally good willing determined by a law that is independent of happiness.[9]

There is another interesting methodological parallel between Kant and Mill: they both try to summon ordinary, pre-philosophical moral judgment in support of their views. Kant frankly acknowledges his lack of revisionist zeal at the beginning of the *Groundwork*. He takes himself to be articulating a principle implicit in everyday morality in that the opening lines about the elevated status of a good will are seen as a tenet of common moral thought. That is why the first section of the book is entitled "Transition from common to philosophical moral rational cognition."[10] From an analysis of ordinary moral conviction we proceed to a philosophically respectable formulation of the principle on which all moral agents tacitly rely.[11]

A few years later, in the *Critique of Practical Reason*, Kant reacts to a reviewer's criticism that the *Groundwork* does not contain a new principle but just a novel "formula" by rejecting the idea of a completely new moral principle as preposterous:

> For who would want to introduce a new principle of all morality and, as it were, first invent it? Just as if before him the world had been ignorant or in thoroughgoing error about what duty is. Whoever knows what a *formula* means to a mathematician, which determines quite precisely what is to be done to execute a task and does not let him miss it, will not take a formula that does this with respect to all duty in general as something that is insignificant and can be dispensed with.[12]

It is obvious that Kant takes his own moral principle, no matter how revolutionary as an exercise in ethical theory, to be firmly grounded in everyday moral thought.

Once again, Mill's approach is very similar. He acknowledges what cannot be denied: there is no agreement among moral theorists regarding the highest principle of morality. Mill brackets the question of how much damage is done in practice by this lack of agreement, but he claims that it would "be easy to show that whatever steadiness or

consistency [the moral beliefs of mankind] have attained, has been mainly due to the tacit influence of a standard not recognized."[13] Of course, Mill's "standard" is the principle of utility.

Does the categorical imperative implicitly rely on utilitarian reasoning?

Yet Mill goes one step further. He argues that the principle of utility – not just considerations of happiness among other things, which would be a much weaker thesis – "has had a large share in forming the moral doctrines even of those who most scornfully reject its authority."[14] He is trying to enlist not only ordinary ethical judgment in support of his cause, and historical figures whose theories he (erroneously) believes to be proto-utilitarian, like Epicurus and the Socrates of Plato's *Protagoras*,[15] but even philosophers whose theories he regards as decidedly anti-utilitarian. According to Mill, even "a priori moralists" have to rely on considerations of utility to make their arguments work, whether they are willing to acknowledge it or not. This includes Kant and his *Groundwork of the Metaphysics of Morals*:[16]

> This remarkable man, whose system of thought will long remain one of the landmarks in the history of philosophical speculation, does, in the treatise in question, lay down an universal first principle as the origin and ground of moral obligation; it is this: – "So act, that the rule on which thou actest would admit of being adopted as a law by all rational beings" . . . But when he begins to deduce from this precept any of the actual duties of morality, he fails, almost grotesquely, to show that there would be any contradiction, any logical (not to say physical) impossibility, in the adoption by all rational beings of the most outrageously immoral rules of conduct. All he shows is that the *consequences* of their universal adoption would be such as no one would choose to incur.[17]

Mill is attacking the core tenet of Kantian ethics. According to Kant, what calls the morality of an action into question is not that it has unfortunate – actual, foreseen, or intended – consequences but that it offends pure reason. This is not to say that it is "irrational" in the sense of modern means–ends rationality (where an action is considered irrational if does not serve the agent's professed purpose). Rather, an action is morally impermissible because it commits the agent to a contradiction when he tries to imagine a world in which the principle or "maxim" underlying the action is universally observed. By its very nature, reason is universal. Reason also abhors contradictions of any kind. This includes the contradiction that would consist in acting on principles that one cannot at the same time want all others to act on as well. It is in this sense that immoral action is contrary to reason.

Mill, however, argues that Kant fails to show that the universality of an immoral maxim directly commits the agent to an inherent contradiction. Rather, the problematic nature of the maxim is said to result from the fact that its universal adoption would have undesirable consequences. We would have to avail ourselves of substantive considerations of value to make the categorical imperative work. If Mill were right, Kant would be some kind of rule utilitarian, not the "a priori moralist" he professes to be. Mill's criticism is spelled out in more detail in chapter 5:

> When Kant (as before remarked) propounds as the fundamental principle of morals, "So act, that thy rule of conduct might be adopted as a law by all rational beings," he virtually acknowledges that the interest of mankind collectively, or at least of mankind indiscriminately, must be in the mind of the agent when conscientiously deciding on the morality of the act. Otherwise he uses words without a meaning: for, that a rule even of utter selfishness could not *possibly* be adopted by all rational beings – that there is any insuperable obstacle in the nature of things to its adoption – cannot be even plausibly maintained. To give any meaning to Kant's principle, the sense put upon it must be, that we ought to shape our conduct by a rule which all rational beings might adopt *with benefit to their collective interest*.[18]

But Mill is mistaken. The interest of *humanity*, collectively or indiscriminately, does not enter Kantian moral reasoning at any stage. What does feed into it is the inevitable *self*-interest of the agent, who takes this to be the natural starting point of deliberation, to see it rebuffed if it turns out that it conflicts with a purely formal moral law.

How the categorical imperative works

Like Mill, Kant assumes that human beings by nature want to be happy: there is *one* end, Kant says, "that can be presupposed as actual" in all human beings, "and thus one purpose that they not merely *can* have, but that one can safely presuppose they one and all actually *do have* according to a natural necessity": their own happiness.[19] This end is composed of the many and varied individual ends that we pursue as human beings. Some of these ends are shared by almost all of us, but many are more personal in that they vary according to the agent's age, gender, culture, occupation, and general outlook on life. But whatever these ends may be, they have in common that seeing them realized gives us pleasure. It makes us happy. Moreover, Kant is committed to the claim that all non-moral action is invariably determined by our overall desire to be as happy as possible.

As we are rational beings with a sense of our own future we would like to see our ends realized overall in the most efficient way possible. In the *Critique of Pure Reason*, Kant emphasizes three distinct dimensions of happiness: we want to see our desires

satisfied "*extensively*, with regard to their manifold nature, as well as *intensively*, with regard to their degree, and also *protensively*, with regard to duration."[20] This is not easily done. That is why we start considering our options when we are prompted by an inclination to act, i.e., we try rationally to assess whether acting on this or that inclination would be beneficial for us.

At some point, however, reason breaks away from prudential deliberation to consider a radically different question: whether the maxim the agent would *like* to act on can be willed as a law to be followed universally, i.e., by all human beings at all times. (Only the Kantian philosopher fully understands how this works, but if Kant is right this is what happens in everyday moral life.) If the proposed maxim survives the test it is permissible to act on it and the agent is free to pursue his interest. If it fails the test it must be rejected, and the opposite maxim must be adopted and enacted instead, purely on the grounds that – unlike the one it was generated from by way of negation – it is the one that *can* be willed as a universal law.[21] The crucial question is what makes a maxim pass or fail the test. If the test can be reconstructed on a formal, purely rational basis, without invoking the "collective interest" of humanity (or any other kind of antecedent value), Kantian ethics is distinct from (rule) utilitarianism and might – *pace* Mill – be a viable alternative as an ethical theory.

Let us briefly examine the duty of beneficence, which may seem closest in spirit to utilitarianism. Like most other moral theorists, Kant thinks that we ought to make other people happy. Doing good is the fourth and final duty derived from the categorical imperative in the *Groundwork*.[22] In the *Metaphysics of Morals* Kant goes so far as to place the field of ethics under the general heading of "one's own perfection and the happiness of others."[23] But how can there be a Kantian duty to make others happy if Kant cannot avail himself of the premise that happiness is good, let alone the highest good?

The case discussed in the *Groundwork* centers on the example of a prosperous man who is conscious of the fact that he could easily share his wealth with those who struggle financially. As characterized by Kant this man is not a vicious person: he does not try to enrich himself by making the poor even poorer. But he is not inclined to share his fortune with his fellow human beings either. He says to himself:

> [W]hat's it to me? May everyone be as happy as heaven wills, or as he can make himself, I shall take nothing away from him, not even envy him; I just do not feel like contributing anything to his well-being, or his assistance in need![24]

His indifference is a reflection of his own natural selfishness.[25] He simply wants to reserve his fortune for his own personal use. But even though this person is not wicked his attitude is still immoral, and so is the careless behavior that comes about as a result. An attitude of indifference is dismissed by practical reason as immoral because the agent

cannot consistently will it as a universal law:

> For a will that resolved upon this would conflict with itself, as many cases can yet come to pass in which one needs the love and compassion of others, and in which, by such a law of nature sprung from his own will, he would rob himself of all hope of the assistance he wishes for himself.[26]

The inconsistency consists in the fact that, his own initial self-assessment notwithstanding, the agent is in fact committed to relying on other people's sticking to a rule he himself is not prepared to adopt. It is a refusal of fair play. The unfairness of the prosperous man does not turn on the value of the actual or intended effects of his actions on human well-being. It merely depends on reason rebutting his own natural selfishness. Inclination, the satisfaction of which makes us happy, speaks first; we can defy it on moral grounds, but we are still committed to its realization. However, as all human beings are essentially equal, endowed with the same dignity, our happiness must not come about as the result of thinking that we are exceptional. We must not rely on others' abiding by the rules we are not ourselves willing to enact. (Similarly, there is a contradiction if you have to rely on the institution of promising to achieve your purpose by means of a false promise, acting on a maxim that if universal would undermine the very institution you are now misusing for your own ends.) Once again the sporting analogy is instructive. If you win the game as a result of taking an unfair advantage you will be relying on your competitor to stick by the rules you are breaking. You would lose your advantage if he were to break the rules too – and at some point you would cease to play the game.

To return to Mill's candidates, the contradiction involved is neither "logical" nor "physical." Nor does Kant have to appeal to the inherent undesirability of certain outcomes. In his ethical theory, what makes an action immoral is that the agent arbitrarily assumes an exceptional status he cannot allow others to share because that might undermine his own success. It is a "practical"[27] contradiction that reason tells us not to incur.

The (not so) common commitment to making people happy

The upshot of this thought experiment is a Kantian duty to make people happy, and a pretty stringent one at that. When we realize that we cannot rationally endorse a maxim we were tempted to act on as a universal law we are rationally required to adopt the opposite maxim instead. A maxim of not helping others cannot be willed as a universal law. The opposite maxim is one of being helpful.[28] As a result we find that we have at least prima facie an obligation to help others whenever we are presented with an opportunity to do so.

It is not difficult to see why Kant also believes that a moral world would be a happy world – by and large, barring natural disasters and other forms of "natural evil." Even if the harm done by lies, thefts, and murders is not the *raison d'être* of their immorality, it is true that lies, thefts, and murders tend to cause much human misery; and a moral world would be free from such evils. In addition, virtuous agents will take their duty to make other people happy very seriously. One of Kant's most enthusiastic endorsements of universal happiness is to be found in Georg Ludwig Collins's copy of notes on Kant's lectures on moral philosophy:

> For God wills the happiness of all human beings, and this by human agency, and if only all human beings together were unanimously willing to promote their happiness, we might make a paradise in Novaya Zemlya.[29] God sets us on a stage where we can make one another happy; it rests entirely upon us.[30]

What is more, happiness is part of the comprehensive good. The more – deserved – happiness, the better.

Because of Kant's insistence that human beings ought to make each other happy, that we must adopt and advance the ends of others as our own, prominent utilitarians and other consequentialist philosophers have tried to explain away the deep theoretical differences between Kant and the utilitarian tradition. In 1993 R. M. Hare famously raised the question whether Kant *could* have been a utilitarian.[31] A few years later David Cummiskey developed his own brand of "Kantian Consequentialism." More recently, Derek Parfit argued that Kantianism, rule utilitarianism, and contractualist theories can be made compatible.[32]

But there are several reasons why Kantian ethics is essentially different from any form of utilitarianism. First, and perhaps most conspicuously, Kant's system relies on a strict distinction between perfect and imperfect duty, which he adapted from the natural law tradition of eighteenth-century rationalism.[33] Roughly, the idea is that perfect duties – such as the prohibition of murder – are negative duties of omission that provide absolute constraints within which imperfect duties – of which beneficence is a prominent example – can become actual obligations. In other words, while the happiness of others is a prominent moral end for both Kantians and utilitarians, the Kantian must deny that there is an actual obligation to promote happiness when it can be done only by violating a perfect duty. The end does not justify the means.

Second, even within the limits set by other moral concerns the Kantian duty of beneficence is not equivalent to the utilitarian master duty to produce the greatest happiness because, for Kant, only the happiness of others has direct moral weight. There is no obligation to make ourselves happy. His official reason for this somewhat surprising thesis is that – as by now we know – all human beings by nature invariably seek their

own happiness. As duty implies restraint, or at least the possibility of tension between what one wants to do and what one ought to do, there can be no duty to promote one's own happiness.[34] By contrast, for utilitarians the theoretical status of the agent's happiness is no different from the happiness of every other morally relevant being. It is a small part of the general human happiness we ought to advance.

Third, utilitarians and Kantians disagree sharply over the criterion for moral status. For utilitarians, a living being is an object of moral concern by virtue of its sentience, i.e., because it can feel pleasure and pain. If our well-being counts, morally, so does theirs. Inflicting pain on an animal is considered wrong *pro tanto* in just the same way as inflicting pain on an adult human being. It is the pain that is bad, regardless of the identity or species of the creature that suffers it. As a result, utilitarians have been at the forefront of the emerging field of animal ethics in the nineteenth and twentieth centuries. For Kant, however, 'animal ethics' is a contradiction in terms, and modern Kantians struggle to count animals in.[35] Kant is not a "speciesist," at least in theory, because he does not restrict moral standing to human beings. A being counts morally by virtue of being a person, i.e., a free rational being that is subject to the moral law. (Mere instrumental rationality is not enough.) A being is a proper object of morality only if it is also a moral subject. The realm of moral obligation is therefore populated by equals, which is a very appealing idea. There are no second-class citizens.

So, if Martians were like human beings in this respect we would have to accord them moral status. It just so happens that human beings are the only kind of person with which we are acquainted. So, de facto only human beings have moral standing in the Kantian system.[36] The welfare of other sentient beings (animals) can be morally relevant, but only by analogy, because of what our actions do to our own character:

> Even gratitude for the long service of an old horse or dog (just as if they were members of the household) belongs *indirectly* to a human being's duty with regard to these animals; considered as a *direct* duty, however, it is always only a duty of the human being to himself.[37]

Kant has to employ this curious strategy to be in a position to argue that cruelty to animals is morally vicious. But utilitarians use similar arguments to explain the moral weight of promise-keeping or to justify the importance we accord to rights. Both utilitarianism and Kantian ethics thus appear to have their blind spots, and both theories use indirect strategies to cope with them.[38]

The fourth difference is the most fundamental one: the utilitarian principle depends on a substantive notion of value whereas Kantian ethics does not. The Kantian agent does not act for the sake of some value, and it is not the production of value that makes moral action right. The categorical imperative enjoins us to do what we have to do because of the *form* of our willing, not because of a desire to achieve something. The Kantian

principle is satisfied only when the agent acts solely from duty, for the sake of conforming to the moral law. Indeed, the moral motive of respect for the law is primarily directed at the act of will, and only by virtue of that at the result one intends to bring about. It is true that a good moral agent will want to promote the happiness of others, but the value of happiness is not the reason why. He wants to make others happy because that is what the moral law commands.[39]

In this respect Kantian ethics is radically anti-consequentialist. Unlike Kant, Mill argues that the rules of action "must take their whole character and colour from the end to which they are subservient."[40] Not so for Kant, for whom the very ends of moral action are first determined by a single, self-imposed rule. So, even if Kant's ethics and some form of utilitarianism were perfectly to coincide at the level of concrete, first-order obligation, (at least) one of the two theories would still be misguided philosophically.

Kantian regard for consequences

So, Kant was not a utilitarian by any stretch of the imagination. His focus on the unconditional goodness of moral willing is incompatible with any recognizably consequentialist framework. Consequences are naturalistic facts that have no moral value; and merely producing facts, no matter how pleasing, has no moral value either.

However, it would be a mistake to conclude that a Kantian agent must ignore the consequences of action altogether (as those who caricature the Kantian system would have it). Consequences matter in at least three different ways. First, as explained above, the foreseeable consequences of the universal adoption of a maxim play a significant role in the thought experiment of the categorical imperative. They are not decisive because the normative question is settled by considering whether the agent can consistently will the maxim of a proposed action, without practical contradiction, as a universal law. The normative work is done by reason alone. But imagining the consequences of a universalized maxim will help us discover whether we can or cannot will a proposed principle as a universal law. This is hardly astonishing given the fact that Kant encourages us to conceive of our maxims as universal laws of nature, i.e., as causal laws describing the behavior of human beings.

Second, the categorical imperative determines the ends of moral action, specified in our maxims, but not the specific means we must take to realize our moral ends. With regard to perfect duties of omission, it is possible that there are no distinct means to be taken to comply with them. I do not have to do anything in particular to refrain from lying when a lying promise seems tempting, e.g., as in Kant's example, to obtain a loan. But some strict duties do require positive activity, e.g., the duty not to break – to keep! – promises. My promise to repay a loan renders it obligatory that I take all necessary steps to produce the money when the time comes. Obviously, the consequences of how I conduct my life are relevant for that purpose. Moreover, calculation of consequences is

even more essential in the sphere of imperfect duty, where the law – as Kant puts it in the *Metaphysics of Morals* – commands "only the maxim of actions," not the actions themselves.[41] We must make it our maxim to be helpful. But having a maxim to make the ends of others one's own is not enough; one must also take the requisite steps to bring it about. Although Kant tends to keep quiet about this, the principles that govern the realization of ends, be they set by the categorical imperative or proposed by human nature, are technical imperatives that prescribe specific means. It is therefore, in this roundabout way, our duty to cultivate our potential to pursue our moral ends skillfully.

Third, the consequences of our actions, even our own happiness, are (indirectly) morally relevant insofar as they affect our capacity to act morally. As we saw above, our own happiness is not something we have a duty to further by itself, but Kant is not blind to the imperfections and limitations of human nature. Misery is hardly a good basis for morally virtuous behavior. When I am unhappy I am much more likely to give in to the temptation to do what is morally bad, e.g., to make a lying promise in order to obtain a loan. As Kant puts it in the *Groundwork*, "lack of contentment with one's condition . . . could easily become a great *temptation to transgress one's duties*."[42] Misery is no excuse for unethical conduct. There is an unconditional duty not to make false promises, no matter how unhappy I am. But that does not entail that I should unnecessarily *risk* unhappiness and the temptations that come with it. Thus making sure that at the very least you are not miserable is indirectly a good thing, morally, after all. But it is still not your happiness that is morally good, but rather your not exposing yourself to moral danger. The same kind of indirect moral reasoning recommends the acquisition of money.[43]

There are many ways in which good consequences enter Kantian practical reasoning even if they do not make actions morally good. But, again, this does not narrow the gap between Kantians and those who consider the value of happy consequences primary.

Conclusion: right action and good willing

Despite the similarity of their origins and their methodological commitments, Kantian ethics and utilitarianism represent two radically different ways of thinking about moral value. Utilitarianism is one of the clearest cases of consequentialism. It starts with an impartial notion of goodness that human actions are meant to bring about. By contrast, Kant's ethics is an example of a much older tradition that starts with the desires of the individual agent. His principle defines the stance human beings should take toward their natural desires, one that can be universally shared. The central practical concept of utilitarianism is right action, that of Kantian ethics the good will.

Perhaps surprisingly, the concept of "right action," which is central to so many debates in contemporary ethics, has no place in Kantian ethics. It is essentially a concept that pertains to the sphere of law or right. In ethics, the notion that comes closest in spirit

to "right action" is action "in accordance" or "in conformity" with duty, discussed at length in the first section of the *Groundwork* as an example of human behavior that is morally lacking. Mere "rightness" has no moral value. For Kant, an action is morally good if and only if it is done from duty, i.e., if it is determined by the moral law and does not just – externally, accidentally – conform with it. He claims that this distinction is rooted in ordinary, pre-philosophical moral thought.

We are now in a position to put one decisive difference between the two types of theory. Despite the fact that both Kant and utilitarians like Bentham and Mill are seeking to identify and establish a supreme principle of morality, their respective favorites play very different roles within their respective ethical systems. For Kant, the categorical imperative must directly determine human action, whereas for utilitarians it is sufficient that human action conform to the principle of utility. But this also means that Kant, but not utilitarianism, is committed to "transparency," in that his theory contains a requirement that the agent must understand – not be deceived about – the nature of morality.[44] Because of the limitations of human calculative capacity utilitarians tend to discourage active reflection on and application of the greatest happiness principle. By contrast, for Kant moral progress can only be achieved by gaining a better understanding of the nature of morality. There is no need, he argues, to teach human beings anything new; we just make them aware, "as Socrates did," of their own principle.[45]

I am indebted to the editors of this volume for their thoughtful comments on earlier versions of this piece, and to Bettina Schöne-Seifert and her colleagues at the Centre for Advanced Study in Bioethics at the University of Münster for their friendship and hospitality in 2012–13.

Notes

1. J. S. Mill, *Utilitarianism*, *Collected Works*, vol. x, p. 205.

2. Kant, *Groundwork of the Metaphysics of Morals*, IV 393.

3. Kant, *Groundwork of the Metaphysics of Morals*, IV 393.

4. Kant does not usually consider the impartial, universal happiness of utilitarianism. The happiness he excludes from the position of the highest good is the happiness of the

agent, which plays a foundational role in eudaimonistic ethical theories. Yet insofar as the goodness of the sum total of happiness relies on the value of individual happy lives, Kant must also reject the idea that maximum overall happiness as such constitutes the highest good.

5. If deserved, happiness is part of the highest good in a different sense: the sum of everything that is good as an end, the "comprehensive" good.

6. Kant, *Groundwork of the Metaphysics of Morals*, IV 393.

7. There is perhaps a slight disanalogy in that the pursuit of happiness need not be competitive. We can all be happy together, deservedly or undeservedly, but fairly or not we cannot all win the marathon at the Olympics.

8. Kant's theory of the (mere) conditional goodness of happiness, even one's own, explains why he would not agree with Sidgwick that there is a 'dualism' of practical reason. For Sidgwick, both egoism and utilitarianism are impeccably rational, i.e., reason cannot decide cases in which self-interest and morality conflict (see the concluding sections of his *Methods of Ethics*). For Kant, self-interest provides reasons for action only if there is no conflict with morality (cf. *Critique of Practical Reason*, v 92–93). The reason is that morality itself emerges as an autonomous restriction of self-interest, not as the result of accepting some independent value outside the agent. We shall return to this topic in the two sections concerning the categorical imperative below.

9. There is also a more subtle difference that divides the two theories. Utilitarianism is committed to the idea that the moral principle is *dependent* on the highest good: that principle is correct which instructs us to bring it about. By contrast, Kant expressly denies that the criterion of what ought to be done depends on the highest good in this way, or indeed on any antecedent notion of value whatsoever. Rather, the supreme good is itself *determined* by the supreme principle of ethics. (Even other goods are good only if they conform to the appropriate standards of reason.) As the categorical imperative is formulated for the first time at the end of the first section of the *Groundwork* we learn that it is this principle that first defines what kind of willing is in itself good (IV 403, cf. IV 437). In other words: for utilitarianism, an action is right if and only if it produces the best possible consequences; for Kant, an action is morally good if it is determined by a principle of pure reason, irrespective of the consequences. It is hoped that the consequences of good actions are good too. But moral and non-moral goodness are sharply distinguished, and happiness is of the latter type.

10. Kant, *Groundwork of the Metaphysics of Morals*, IV 393.

11. A bit further down, to explicate the concept of an unconditionally good will, "as it already dwells in natural sound understanding and needs not so much to be taught as rather just to be brought to light" (IV 397), Kant proposes an analysis of the concept of duty, which will lead to the formulation of the supreme principle of morality at IV 402. Summing up the results of the first section of the *Groundwork*, he says that there is no need to teach us anything new. We just have to make the human mind, "as Socrates did," aware of its own principle (IV 404).

12. Kant, *Critique of Practical Reason*, v 8, fn.

13. J. S. Mill, *Utilitarianism*, *Collected Works*, vol. x, p. 207.

14. J. S. Mill, *Utilitarianism*, *Collected Works*, vol. x, p. 207.

15. There is a question as to why Mill regards the hedonists of antiquity as utilitarian, given that their hedonism was deliberately and openly egoistic: Epicurus held the view that actions are good insofar as they promote *the agent's* happiness, not – like utilitarianism – the happiness of all considered impartially. Did the Enlightenment ideal of impartiality have such a powerful influence on Mill that ancient individualism was beyond the powers of his imagination?

16. Mill omits the word *Grundlegung* (variously translated as 'Foundations', 'Principles', or, in recent years, 'Groundwork') from the title of the work he calls the "Metaphysics of Ethics," but he is clearly referring to the *Groundwork* of 1785, not the *Metaphysics of Morals* of 1797.

17. J. S. Mill, *Utilitarianism*, *Collected Works*, vol. x, p. 207.

18. J. S. Mill, *Utilitarianism*, *Collected Works*, vol. x, p. 249.

19. Kant, *Groundwork of the Metaphysics of Morals*, IV 415. Mill, in chapter 4 of *Utilitarianism*, also takes self-interest as his starting point, and proceeds to make it compatible with utilitarianism by arguing that the ends of others should be incorporated into one's own. By contrast, Kant assigns pure reason the task of rejecting self-interest as a determining feature of the human will altogether, not in favor of collective interest but

rather in favor of adhering to a pure law of reason for its own sake.

20. Kant, *Critique of Pure Reason*, A 806/B 834.

21. This is the maxim that was generated by negating the one suggested by self-interested deliberation. Pure moral judgment is not driven by interest; on the contrary, pure judgment itself generates an interest in acting morally – if necessary contrary to all pre-existent desires – called "respect for the law."

22. Kant, *Groundwork of the Metaphysics of Morals*, IV 423; cf. IV 430.

23. Kant, *The Metaphysics of* Morals, VI 385. Note that (1) this excludes the sphere of perfect duties of right (juridical duty) and that (2) utilitarians would, of course, reject the *perfection* of the agent as morally relevant, and other purely self-regarding duties along with it.

24. Kant, *Groundwork of the Metaphysics of Morals*, IV 423.

25. The fact that Kant has to rely on natural pre-moral selfishness to make the categorical imperative work is frequently overlooked. Agents do not try to "universalize" any old rule that enters their head but maxims they deem advantageous on the basis of weighing their inclinations. Without this input the categorical imperative would not yield conclusive results. It *would* be empty. But note that this does not mean that Kant's actual ethical theory is egoistic. On the contrary, our natural egoism is pushed back by the moral law. If the indifferent agent ultimately realizes that he ought to help and on that basis decides to do so, it is simply because he knows beneficence to be his duty, not because of a strategic decision, e.g., the hope that the other might one day come to his rescue.

26. Kant, *Groundwork of the Metaphysics of Morals*, IV 423.

27. See Korsgaard, "Kant's Formula of Universal Law" for an overview of possible types of contradiction, and a defense of the "practical contradiction" interpretation.

28. The stringency of imperfect duty is controversial among Kantian ethicists. Thomas Hill's "Kant on Imperfect Duty and Supererogation" is the most prominent defense of a less exacting interpretation of Kant's theory of imperfect duty, according to which the

mandatory maxim is merely that of helping *sometimes*. But any such restriction of the principle of beneficence would be arbitrary. As Kant puts it in the *Groundwork*, we have a duty to be beneficent where or whenever we can (IV 398). I defend a more demanding reading in "Good but Not Required?" See also Marcia Baron's classic *Kantian Ethics Almost without Apology*, particularly part I.

29. A proverbially bleak Russian Arctic archipelago, used as a nuclear test site during the Soviet era.

30. Kant, *Lectures on Ethics*, XXVII 285–286.

31. Cf. my late reply in *Utilitas*, "Why Kant Could not Have Been a Utilitarian."

32. Parfit, *On What Matters*, vol. I, chapters 16–17.

33. For a historical discussion of how the distinction emerged see J. B. Schneewind's "The Misfortunes of Virtue."

34. Cf. *Metaphysics of Morals*, VI 386. The happiness of the agent is morally relevant at most indirectly, e.g., when there is a temptation to neglect one's own well-being so much as to make committing a crime irresistible. See the section "Kantian Regard for Consequences" below.

35. See Christine Korsgaard's *The Sources of Normativity*, especially chapter 4, and my own "When the Tail Wags the Dog" for two very different attempts to remedy the situation.

36. I bracket the case of God. According to Kant, there is reason to believe in a personal God on moral grounds, who must be regarded as the head of the moral community of rational beings. But as we are not acquainted with God there can be no duties to God. See *Metaphysics of Morals*, VI 443–444.

37. Kant, *The Metaphysics of Morals*, VI 443.

38. While the inclusion of animals may appear to favor utilitarianism it also contains the seed of a well-known objection to it. Utilitarians are often accused of not caring about individuals, which are regarded merely in their capacity as "receptacles of pleasure." In

Kantian ethics, moral status seems to confer a dignity utilitarianism is unable to account for.

39. The most striking formulation of this idea can be found in Kant's rejection of heteronomy at the end of the second section of the *Groundwork*: "Thus I ought e.g. to try to advance the happiness of others, not as if its existence made any difference to me (whether because of immediate inclination, or some delight indirectly through reason), but merely because the maxim that excludes it cannot be comprised in one and the same willing, as universal law" (IV 441).

40. J. S. Mill, *Utilitarianism*, *Collected Works*, vol. x, p. 206.

41. Kant, *The Metaphysics of Morals*, VI 390.

42. Kant, *Groundwork of the Metaphysics of Morals*, IV 399.

43. Kant, *The Metaphysics of Morals*, VI 388.

44. Cf. B. Williams, *Ethics and the Limits of Philosophy*, p. 101.

45. Kant, *Groundwork of the Metaphysics of Morals*, IV 404.

13 What virtue ethics can learn from utilitarianism

Daniel C. Russell

I am not a utilitarian. My own thinking in ethics has been in line with a very different tradition, one in which the central concept is an excellent character trait – a virtue, for short – and so that tradition is nowadays called virtue ethics. Utilitarians say that whether an action or policy is right or not depends entirely on its consequences for all concerned.[1] Virtue ethicists say something very different: ultimately what is right is to be a certain kind of person, a person with the virtues of character – generosity, for instance, and fairness, and benevolence – which express themselves not only in feeling and motivation but also in action.[2]

The obvious thing for me to do in this chapter, then, would be to focus on the differences between these two traditions, and especially on why I think those differences speak more in favor of virtue ethics than utilitarianism. So it may come as a surprise that that is not in fact what I am going to do: I come not to bury utilitarianism but to praise it. Or, if not strictly that, then at least to say what I think virtue ethicists like me can learn from the sorts of cause-and-effect thinking that have always been the hallmark of good utilitarian philosophy.

Why everyone has to think about consequences

Taking consequences seriously is not the same as consequentialism

We should clear something up right away: there is consequentialist thinking, and there is thinking about consequences. The two are different. What I mean by consequentialist thinking is the view that whether a private action or public policy is right depends on the consequences of that action or policy, *and on nothing else*.[3] In other words, it is the sort of ethical thinking that characterizes "consequentialism" as a distinctive philosophical approach in normative ethics, of which utilitarianism is the main species. Thinking about consequences, on the other hand, is simply that: it is to acknowledge that our actions have consequences we have to consider in order to choose our actions responsibly. That is so whether or not we accept the specifically consequentialist thought that consequences are *all* there is to consider.

Although the difference between these two kinds of thinking is clear enough, it is not uncommon to hear any argument that draws attention to the consequences of some action or policy described as a "consequentialist argument."[4] Unfortunately, since many

philosophers also think – mistakenly, in my view – that consequentialism takes the moral "low ground," this slide can make it look like we have to choose between taking the high ground and really taking consequences seriously ("Well, yes, of course consequences should be considered, but let's not lose sight of what *really* matters here . . ."). We have to avoid this slide, because if we have to choose between being virtuous people and taking consequences seriously, we are doomed.

That slide also blinds us to the importance that consequences have always had in all traditions in ethics. For example, if we give in to the temptation to divide philosophical schools by crisp, stark contrasts, we might say that Jeremy Bentham is the one who thinks morality is simply about weighing costs and benefits, say, whereas Immanuel Kant is the one who thinks that what matters is the motive.[5] That would get both of those thinkers importantly wrong, though.

The mistake is not, of course, the thought that Kant *does* focus on motives. The mistake is in skipping over the fact that Kant also thinks one of the chief things we should be motivated to do is to create a world in which we all have better opportunities to live happy lives and enjoy the things we want.[6] In that case, part of what it *is* to have the right motive is to consider the consequences of our actions and policies, asking whether they will move us closer to that kind of world or farther away from it. That is why Kant thinks that to have the virtues is to have the goals both of developing ourselves in ways that make us useful to each other and of choosing actions and policies that will make people better off.

Furthermore, even though Bentham develops a radically different approach from Kant's, still he argues for the "principle of utility" mainly on the grounds that we owe it to each other to have public policies that actually serve the public good. What motivates Bentham is his outrage over corrupt systems allowing policy-makers to help themselves and their cronies at public expense, exempting themselves from principles of fair play they expect of everyone else.[7] His proposal is that public policy should instead be shaped in a principled way, without prejudice, giving fair and equal consideration to everyone affected, and taking results seriously.[8] To be sure, Kant did not agree with Bentham that the principle of utility was the right principle to live by, but they both agreed that we ought to try to make each other better off, and that there is not much that is good about our intentions if we do not think carefully about what would happen if we acted on them.

For both Bentham and Kant, then, we do not take consequences seriously in spite of the fact that people are what really matter, but precisely *because* people matter. And it is no different for virtue ethics.

How consequences matter in virtue ethics

Thinking about consequences is an important part of virtue ethics because part of acting virtuously is aiming at certain consequences and not others. Take for instance the virtue

of generosity. As Aristotle puts it, to say that someone has a virtue, such as generosity, is to say that he or she has the right goal,[9] namely to help others by sharing resources with them.[10] Now, already it is clear that a generous person will need instrumental reasoning about what resources are needed, what resources are available, efficient ways of using them, and so on. The goal of generosity is to do things that *actually help*, so results count. In fact, Aristotle says that the people who really have virtues of character are the ones who back up their good intentions with virtues of practical intellect, including instrumental savvy and know-how.[11]

As important as that idea is, though, there must be more to thinking about consequences than that. If thinking virtuously about consequences were strictly about finding effective means to virtuous goals, then that would mean that we could know exactly what goals are virtuous in *advance* of thinking about the consequences of different ways of pursuing those goals. To see why that is not so, consider what happened in Colombia in 2003, when a tribe of refugees emerged from their traditional home, displaced by guerilla fighting. Aid workers responded to their plight by setting up a camp and guaranteeing a steady stream of provisions. The refugees survived their ordeal, but in the long run living in the modern world for them came to mean waiting for the next supply truck to arrive. Before long, the aid workers themselves realized that the help they gave was not really the help the refugees needed.[12]

What went wrong? Not the broad goal – the aid workers really were trying to help. Not the means, either: it is not as if they collected the wrong supplies or sent them to the wrong camp. What, then? Notice that although "helping refugees" is a generous goal, it is also an *indeterminate* one. Consider an analogy: healing is a doctor's goal, but only in the very broadest, abstract sense; it becomes a goal that a doctor *can actually pursue* with a given patient only when the doctor determines what healing would *amount to* in that patient.[13] It is with this middle step – making indeterminate goals determinate – that things went wrong in the efforts to aid the refugees. The problem was not the goal or the means, as far as those went, but a failure to understand what *genuinely* realizing that goal, then and there, would look like. That is why three years later one aid worker said in retrospect, "People want to protect [the refugees]. To help them, we give them food and clothes. That doesn't help them at all in the long term."

Aristotle calls the ability to perform this middle step well *phronēsis* – practical intelligence or wisdom – and he says that it is always guided by a deeper understanding of what it takes for people to live well.[14] That virtue of practical intellect picks up where plain common sense and gut instincts about what "feels generous" or "looks fair" leave off, where we need a more intelligent and deliberative appreciation of what role generosity and fairness have to play in our lives, what goods they help us realize, and what problems they help us solve.[15]

Practical intelligence about consequences

Practical intelligence helps us act virtuously in a couple of specific ways, each of which involves very careful thinking about consequences. The first should be fairly clear already. The virtues each have certain characteristic goals: helping others in the case of generosity; in the case of courage, protecting what is valuable even when it is risky to do so; in the case of justice, doing what is equitable; and so on.[16] Such goals are results, so practical intelligence must make those goals determinate by understanding not only what matters but also what really works. That means there is no knowing what the generous or courageous or just thing to do would even be in advance of thinking about consequences.

Practical intelligence also helps us act virtuously in an *overall* way.[17] A virtue gives us the right goal, but not always the same goal as other virtues, so often the best we can do is to strike the best balance we can.[18] For instance, generous people wish to share with others, but if they are also just, then they will not share what has already been promised to someone else. More generally, we have to balance our goals whenever resources and opportunities are limited – which is to say, all the time. Money spent on defibrillators for the hospital (one virtuous goal) is money unavailable to be spent on books for the library (another virtuous goal), and it takes practical intelligence to strike a wise balance.[19] The cost of pursuing any virtuous goal is the forgone opportunity to pursue a different virtuous goal, so virtuous people have to think about consequences like costs and benefits.

We can see all of this more clearly by concentrating on several different kinds of cases where practical intelligence involves careful thought about consequences.

Hard cases and dilemmas: their treatment and prevention

Easy cases and hard cases

Is practically intelligent thinking about consequences really as important as I have said? After all, sometimes what virtue requires is just obvious. Consider this case:

> Harriet has just learned that her aged mother has been rushed to the hospital and will die without heroic life-saving measures. After consulting with several physicians to understand all the relevant facts about her mother's situation, Harriet learns that even with heroic measures her mother's prospects would be horrendously painful and debilitating. Harriet can either approve or deny heroic measures for her mother.[20]

This is an easy case – not because Harriet will find it easy to let her mother die, of course, but because it is easy to know what Harriet should do. It does not require practical intelligence to figure that out, just basic decency and common sense.[21]

Michael Slote argues that this is actually how things are all the time for truly virtuous people. Even though a case like Harriet's requires carefully gathering a lot of complicated facts, Slote says that once the facts are in, a truly virtuous person will always find it obvious what to do, without deliberating. If Slote is right, then perhaps virtue does not require practical intelligence – a virtue of deliberative reasoning – after all, much less deliberation about consequences.

I disagree. I think that there really are hard cases, cases where being virtuous and having all of the facts still does not make it obvious what to do. But even more than that, it takes practical intelligence and often thought about consequences to know when a case actually *is* easy in the first place. Consider this case:

> In order to improve air quality, we are given a choice between two policies requiring car manufacturers to reduce emissions in new automobiles. One of the policies calls for more ambitious reductions than the other.

This looks like an easy case with an obvious answer: do we want a bigger improvement in air quality or a smaller one? But that is to assume that we are being given a choice between two *outcomes*, when in fact we are given a choice between two *policies*, and it takes practically intelligent thinking about consequences to know which policy *will in fact* do more to improve air quality (and at what cost).[22] Here is why. Suppose that the more ambitious policy would make new cars more expensive than the less ambitious policy would. Rising prices discourage consumption, so under the more ambitious policy more people might decide to keep their old cars longer, resulting in *more* total pollution than if we had adopted the *less* ambitious policy. If so, then the more ambitious policy might show that our hearts are in the right place, but at the cost of more pollution than there had to be.[23] This case, then, is actually a hard one. Evidently, it can take a lot more thinking than it may seem – including thinking about consequences – to say whether a case is hard or easy in the first place.

Similarly, in retrospect we see that Colombian aid workers faced a *hard* question in 2003, but at the time we too probably would have thought the question was simply *whether* we really wanted to help or not. That question is easy, but the road to unintended consequences is paved with easy answers to hard questions. In cases like that, choosing the "obvious" solution can itself become the problem. Some cases really are easy, but the danger is that we might think we see easy cases even when we are actually looking at hard cases. Knowing what to do in cases like those is very difficult indeed, and it takes practically intelligent thinking about the consequences of our decisions not only to know what to do, but even to appreciate why knowing what to do

in such cases *is* difficult in the first place.

Dilemmas and "the right thing to do"

Dilemmas are special sorts of hard cases in which all of our options are terrible. Consider the following example. During one of the 2011 debates leading up to the US presidential election, one of the candidates was asked this hypothetical question:

> A healthy 30-year-old young man has a good job, makes a good living, but decides, "You know what? I'm not going to spend $200 or $300 a month for health insurance because I'm healthy, I don't need it." But something terrible happens, all of a sudden he needs it. Who's going to pay if he goes into a coma, for example? Who pays for that?[24]

That is an important question and a presidential hopeful should have something intelligent to say about it. However, notice also how tempting it is to think that having an intelligent answer means saying what would be a solution we can find fully satisfactory – something that would be "the right thing to do." The problem with that tempting thought, though, is that every option in cases like this one is cause for regret, not satisfaction, from the perspective of the virtues.[25] The more we shift the cost of expensive treatment toward the patient, the more we may risk ruining his finances; the less we shift it his way, the more we put the cost of his decision on someone else. Being merciful, we would regret doing the former, and being fair, we would regret doing the latter.

Of course, this does not change the fact that we have to do *something*, or the fact that one of the things we might do could be *superior*, even *vastly* superior, to the other. Likewise, it is still possible to make one's choice thoughtfully, intelligently, and benevolently, as a virtuous person would. But the point, for virtue ethics, is that sometimes we just have to choose what to regret, even when we choose well. Sometimes it is built into the situation that the most excellent choice we can make cannot be better than choosing the least of evils.[26] That is cold comfort, I know, but it does spare us the illusion that deliberation can reach an appropriate conclusion only when it concludes with a genuinely satisfying choice. If instead we suppose that there just *has* to be such a thing as the right thing to do, we shall probably go in circles seeking and not finding it – which, too often, is exactly what such debates look like in philosophy. If nothing seems really satisfactory in a dilemma like this, it is for good reason: from the perspective of the virtues, nothing is.

What practical intelligence does in a dilemma

So how does practical intelligence help us find our way out of jams like this one? There is not a simple answer – it is the nature of dilemmas, after all, to be anything but simple.

But what is important for us to see at the moment is that practical intelligence has to reckon with consequences here, in a number of ways. For one, it takes practical intelligence to tell when one really *is* in a dilemma in the first place,[27] and often one cannot tell without thinking about the consequences of different options. As we saw above, so here too things can look simpler than they actually are. For instance, perhaps it will seem that the thing to do about the uninsured patient really is simple: mercy is a virtue, so the right thing is to simply forgive his debt; true, this will also make health care more expensive, but that just means that the burden will be spread among those better able to bear it. But what if forgiving his debt makes health insurance more expensive for those better able only because it makes it more expensive for *everyone*? Would subsidizing risky choices mean we could expect a lot more of them to be made? On the other hand, sometimes things are actually simpler than they seem: for instance, perhaps we have not considered ways of structuring the patient's debt so that his risk really costs him something without ruining him. The point is that a dilemma exists whenever all of our options are regrettable from the point of view of the virtues. It takes practical intelligence to know when that is the case, but usually that is not knowable apart from the consequences of the options we have.

Furthermore, because dilemmas force us to choose what to regret, we need practical intelligence to know, for instance, when we would lose more of the things we value than we would gain – in a word, how much our various options cost. This is not to say that there is some "common currency" by which costs are measured. There is no common currency between regrets from the perspective of fairness and regrets from the perspective of mercy, and the fact that practical intelligence is ultimately guided by an understanding of human well-being does not change that. But again, dilemmas force us to choose, and when forced to choose we can, even when values are incommensurate.[28] Making that choice virtuously means knowing what we were sacrificing – that is why in dilemmas it is regret, not satisfaction, that virtuous people feel.

To say that costs matter is not to give cost–benefit analysis the last word, though. David Schmidtz puts the point this way:

> If enacting a certain proposal would help some people and hurt others, then showing that winners are gaining more than losers are losing counts for something, but it is not decisive. One must then argue that the gain is so great for some people that it justifies imposing a loss on other people. In contrast, to show that losers are losing more than winners are gaining should pretty much end the conversation.[29]

To show that benefits exceed costs is significant but not decisive, Schmidtz says, because "not all situations call on us to maximize what is valuable. Promoting value is not always the best way of respecting it. There are times when morality calls on us not to maximize value but simply to respect it." Thinking about consequences is one of the things that go into choosing virtuously; but it is not the only one, and it takes practical intelligence to

know when it is decisive and when it is not. But because consequences *do* have the potential to end conversations, practical intelligence must consider consequences even when it goes on to decide that maximizing what is valuable is not the best way of respecting it.

What practical intelligence does about dilemmas

Something else to notice about the uninsured-patient case is that by focusing on addressing a problem that has already arisen, it diverts attention away from why that problem ever arose at all.[30] This brings us to what may be the most important way that practical intelligence deals with dilemmas: shifting from the level of one-off dilemmas to the level of making dilemmas a much less frequent part of our lives in the first place. Think of it this way: if a dilemma is a case in which all options are regrettable from the point of view of the virtues, then from that perspective the most virtuous thing to do is to keep things from ever coming to the point of a dilemma, if we can.

Here we shift from thinking of what practical intelligence does *in* dilemmas to what it can do *about* them. That is also a shift from thinking of the consequences of dilemmas to thinking of dilemmas as the consequences of other things – decisions we might stop making, or failures to take steps that would have left us with nothing to regret. Because questions like what to do about uninsured patients are important ones, it is *all the more important* to ask: Why are there patients without insurance in the first place? What trail of policy decisions has led us to such dilemmas? What policy decisions would be more responsible? What would really work? Might an ounce of prevention actually be worth much more than a pound of cure?

Let's think about prevention. One way to take care of a car is to wait until something fails and then take the car to a mechanic. This is always inconvenient and usually very expensive. Another way is to make regular visits to a mechanic for preventative maintenance. This will not make emergency visits entirely redundant, unfortunately, but it will make mechanical failures a lot less likely to happen and a lot less severe when they do happen. The first approach is necessary once a mechanical failure has occurred. The second is necessary because there is an important extent to which mechanical failure is in our power to avoid.

Exactly the same is true about dilemmas. It takes practical intelligence to sort out dilemmas once they have occurred, and it takes practical intelligence to know how to make dilemmas a lot less likely in the first place. The dilemma of the uninsured patient involves an obvious medical emergency, but it is also a kind of *social* emergency – someone turns up needing urgent care that he is not prepared to pay for. Question: Why don't sick people *always* find themselves in that sort of emergency? The answer is that we have social institutions (whether public or private) that give us better options than waiting for an emergency and then hoping for the best. Following the analogy with preventative maintenance, we can call this "preventative problem-solving." Because

preventative problem-solving is so important for virtue ethics, I want to focus on it in the rest of this chapter.

Taking responsibility for consequences: the institutional approach

When we talk about preventative problem-solving, we are usually talking about social institutions. One of the chief ways that practical intelligence helps us do the right thing about dilemmas, then, is through good institutions. Without them, the everyday challenges of feeding, clothing, and sheltering ourselves become daily dilemmas. With them, we have better things to do with our time than coping with dilemmas – and a lot less to regret. That, I think, is an excellent way to think about social institutions: they are tools of preventative problem-solving.

Emergency measures are not institutional solutions

It is surprising that this way of thinking about institutions is not more common than it is. Consider this case, adapted from G. A. Cohen:[31]

> Two people, Able and Infirm, each need the best chance of thriving on a desert island. (a) The island's *resources* could be divided between them. Even with equal shares, Infirm may not be able to produce enough goods from his share of resources to survive. (b) *Authority* over the island's resources could be divided between them. Infirm could permit Able to use all of the island's resources he wants in return for an agreed-upon share of whatever Able produces (say, 50 percent).

The choice, then, is between (a) private ownership and (b) joint ownership, and Cohen thinks that (b) is fairer than (a). I agree, but there are two things we need to note about this case.

First, this case *is* a dilemma: each option involves something regrettable from the point of view of the virtues, namely the fact that someone will not really have a life of his own. On the first horn of this dilemma, Infirm will be eternally beholden to Able for his survival; on the second, Able will be eternally beholden to Infirm for permission to be productive.[32] The fact that the second horn may be superior to the first does not make it any less a sorry compromise.

Second, and more important, it is the peculiar nature of two-person desert island life that explains why Able and Infirm have this dilemma in the first place. Social institutions are supposed to give us all less to regret by making it possible to improve our prospects in ways that are not zero-sum. Consider that for much of human history simple near-sightedness has been enough to make someone an Infirm. With the right institutions,

though, production and exchange have created an industry of vision correction that turns Infirms into Ables by the millions, and not at the *cost* of people in that industry getting on with their own lives but as a *way* of getting on with them. The problem in the desert island case, though, is that such institutions inevitably remain beyond the pale, so even though that case may show what justice requires in certain sorts of emergencies, we are still left asking what justice requires in terms of institutions for making those emergencies a thing of the past.

What is virtuous about institutions?

So let's look at a case that distinguishes emergency plans from institutional solutions to focus on the latter:

> A large group of people arrive on a desert island. Their plan for early survival is that everything they produce should go into a common stock from which each of them should have all of his provisions. They can then continue with this plan, letting it frame the basic institutions of their community, or they can change to different institutions.

Except for the bit about the desert island, this is actually the case of the early years of the Plymouth Colony in North America. Before setting sail, the colonists drew up a charter in the summer of 1620 that provided for taking from each according to his ability and giving to each according to his need, and I think this was an excellent plan for coping with their impending emergency. Unfortunately, that plan was meant to last for seven years, but within three years the plan had itself become the problem: it meant that the person who sowed would not be the one to reap, and this both discouraged productivity and bred resentment and unrest. The colonists' new plan was to divide their land into privately owned parcels so that sowers would also be reapers, and a rush to consume was quickly replaced by a rush to produce. This meant not only greater prosperity for everyone but also a much more peaceful and cohesive community.[33]

The colonists undid the consequences of one set of institutions by replacing them with institutions that worked much better for them. It is a plan that can warm a utilitarian's heart, but shouldn't it leave a virtue ethicist cold? Shouldn't the colonists have put more of their energy into being more virtuous than into creating cold, commercial institutions? For instance, Michael Sandel has recently lamented "the corrosive tendency of markets," arguing that doing things for commercial reasons means not doing them out of virtue, and as Aristotle observed, virtue is something that grows only with exercise.[34] In fact, the idea that commercial institutions make us less reliant on the virtue of others goes back to Adam Smith himself:

> It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest. We address ourselves,

> not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantages.[35]

Doesn't that make such institutions themselves regrettable from the perspective of virtue ethics?

It is easy to suppose that institutions like private property and markets would be redundant if everyone were virtuous, but that is not so. Even virtuous people need to know what belongs to whom, as well as a peaceful system for settling questions of what belongs to whom. Not just ruthless corporate banks but even the Self-Employed Women's Association need it understood that they are the owners of their capital, so that they can continue making life-changing loans to poor Indian women. Likewise, institutions like a price system provide not just the vicious but also the virtuous with crucial information about when the resources they are using are becoming scarce, in what lines of work their labor is most needed, and so on.[36]

However, there is an even more fundamental point here: the implicit assumption that it is virtuous to sow without expecting to reap is in fact false. Of course, it is virtuous to be prepared to make sacrifices out of what one produces, but that is not the same as being prepared to live in a system where production itself is a continual act of sacrifice. Even virtuous persons are entitled to say, "I have my own life to live, and this just costs too much of my time." Opportunity costs count, and so it is virtuous to have institutions that compensate people for the forgone opportunities that their productive activities cost them.

Perhaps that is why Aristotle himself, who does believe that virtue grows strong only with exercise, also believes that the sorts of institutions eventually adopted at Plymouth Colony make for stronger communities. In large groups of people, Aristotle says, it is concord, not quarreling, that requires explanation, and concord is even rarer in groups that have everything in common than in groups with private ownership, even if the first groups are smaller. The reason for this, he says, is that we naturally take joy in having things of our own and sharing generously, and with having everything in common we get neither.[37] Aristotle does not say anything about the need for private ownership and exchange as a concession to vice, but vice is not the point anyway. In framing institutions it is virtuous to be realistic about the fact that not everyone is; by contrast, there is nothing virtuous about imposing huge costs on people and hoping that virtue will somehow make up for that.

From Aristotle's point of view, we should say that there certainly was a failure of virtue in the case of Plymouth Colony. However, it was not the failure of the colonists to be more virtuous under a fragile system. It was their failure to have a less fragile system in the first place.

If it is just business, where is the virtue?

The argument that commercial society crowds out virtues makes another false assumption as well: it assumes that because market transactions are motivated by commercial interests, market society does not call on participants to be virtuous. It is true of course that the baker gives me bread not out of benevolence but because I pay his price. Where, then, is the virtue? At the very least, it is in the fact that I do not just take the bread. But more than that, it is in the respect of not only baker and customer but all citizens for the institutions that make such transactions possible.

Why are those institutions worth respecting? One reason is that they create not just prosperity, but prosperity that spreads everywhere. Plato offers a case to illustrate this point:

> A community of four people is to produce four essential goods: food, shelter, clothes, and shoes. (a) Each person might produce all four goods for himself. (b) Alternatively, each person might produce one good and exchange for other goods.[38]

Although Plato says famously that philosopher-rulers will have "all things in common," he also says that productivity comes only from a system of exchange, option (b). The more that someone can focus on one task, the more skilled he can be, the more time he can save, and the more he can produce. (For that reason, it is only a system of exchange that makes it possible for there to be a group with the luxury of having all things in common.) What is more, the benefits of such a system quickly spread: not only do goods improve in quality, availability, and affordability, but there are also more opportunities, and more diverse opportunities, for people to make a living. Someone who might have starved if he had had to grow his own food, Plato observes, now can make a living by selling a farmer's produce in the market-place. Not only does such a system completely redefine who is "Able" and who is "Infirm," it also makes it increasingly less necessary for *anyone* to play the part of "Infirm" in the first place, by increasing the ways in which anyone might become an "Able."

Production is one side of prosperity, and the other side is conservation. Institutions can require the vicious to conserve, but they can also give the virtuous the power to conserve. Suppose that a group of farmers graze their cattle on a pasture that none of them owns. At some point adding one more cow to graze will drop every cow's share below the adequate level. When one of the farmers adds that extra cow, every other farmer must either add no more cattle to the pasture and sustain a net loss, or add more cattle in hopes of extracting what value from the pasture he can before it is all used up. As the risk increases that other farmers will do the latter, there ensues a race to deplete the pasture.[39] Now, if we say that those farmers should be more virtuous and save the pasture, we miss the point that even the virtuous farmers among them are powerless to make the sustainable living that they, *being virtuous*, most strongly prefer. What they

need is not more virtue but more power behind their virtue. Institutions like private ownership empower conservers to limit access to exhaustible resources – for instance, by pricing them in such a way that depletion becomes too expensive.

A further kind of reason to respect those institutions is that they enable us to respect each other as we get the things we need from each other. When I buy bread from the baker, we both acknowledge that neither of us is forced to do business with the other. Because he does not have to sell me any bread, and because I do not have to buy bread from him, each of us has to give the other a reason for there to be any transaction between us at all. And in fact, I think that was Smith's point all along: transactions happen only when someone makes the first move, and the decent first move to make is not a demand but an offer. That is why I pay the baker's price instead of saying, "If you were virtuous, you would give me a loaf of bread." As Smith says in the very same passage, "Whoever offers to another a bargain of any kind, proposes to do this: Give me that which I want, and you shall have this which you want, is the meaning of every such offer." By transacting by the rules of commerce, we each get what we want *and* show each other respect in the process.

Institutions like these also enable us to show respect for each other by letting things that should be just business actually be just business. Personal virtue can require that we let some things be impersonal. For example, Michael Sandel's book, *What Money Can't Buy*, is one of the things that money can buy. Its publisher sells it for a profit, along with even more titles in areas like marketing, finance, and business. This for-profit venture presumably has no *intrinsic* desire to publish a book on the corrupting power of profits. But the good news for Sandel is that in market society, the publisher can have a reason to publish that book even *without* an intrinsic desire for it. Likewise, the good news for a reader like me is that the only thing standing between me and a copy of that book is the price. I do not also have to convince my neighbors that I should be allowed to buy it, any more than Sandel had to convince his neighbors that he should be allowed to sell it. When institutions are as impersonal as that, we do not have to agree on who is correct about what should or should not be for sale. We only have to agree to leave that decision with the buyers and the sellers. Of course, that means that buyers and sellers may well make choices that you and I will find repugnant, even unconscionable. But as John Stuart Mill observed in *On Liberty*, the desire to follow one's conscience and the desire to keep others from following their conscience are not equal: one is a desire to be free and the other a desire to deprive others of similar freedom, exempting oneself from "reciprocity of obligation towards them."[40] Sandel is right to say that markets are cold and impersonal[41] – but I say, thank heavens for that. As Gerald Gaus says, "minding one's own business is a real virtue in a free and pluralistic society."[42]

So, even though *each trade* within commercial institutions is an act from commercial incentives, not virtue, nonetheless *participating in that system of trade* can itself be the virtuous thing to do. Participating in that system is one way for a virtuous person to

recognize the right of others to live their own lives, and to take responsibility for meeting his own and others' needs through a scheme that increases available goods.

An idea that is emerging from this discussion is that it is a mistake to suppose that we can say anything determinate about what virtues like justice or benevolence require in advance of knowing what institutions are actually feasible. What virtue requires depends on what actually works. I want to explore this idea a little further now in the last part of this chapter.

Feasibility first

Tough on vice, strong on virtue

Let's look at another case:

> In 1996, a hurricane struck the Raleigh-Durham area of North Carolina. In the ensuing power outage, ice became extremely scarce. Four young men in Greensboro, North Carolina bought 500 bags of ice for $1.70 a bag and took them in a refrigerated truck to the afflicted area. There they sold the ice for $12 a bag, before being arrested for price-gouging.[43]

When ice is $12 a bag, people who want ice to chill their beer might choose to go without, but others who need ice to chill their insulin (say) have little choice but to pay. Like many people, Michael Sandel believes that price-gouging ought to be illegal: greed is at odds with civic virtue, and even if a good society cannot banish greed, it "can at least restrain its most brazen expression, and signal society's disapproval of it."[44] If it is obvious what civic virtue requires, Sandel suggests, then virtue should come first and public policy should follow. Anti-gouging laws do that by showing that we are both tough on vice and strong on virtue.

Do those laws also get ice to the people who need it most? Let's check some figures. Among US states with anti-gouging laws, the most lenient (Utah) allows casual sellers a 30-percent mark-up. A 30-percent mark-up on $1.70 adds $.51 to the price of each bag, or $255 on 500 bags. From that $255 we must subtract the cost of renting refrigerated trucks, the cost of fuel, food, and accommodation (at scarcity prices), not to mention things like cutting and clearing fallen trees from the road, and of course the sellers' lost time. Now, at the more common 10-percent legal mark-up, these costs would have to be subtracted from just $85. But even the most lenient legal standard probably would have required selling at a loss. A more attractive option is not to sell at all, which of course is always legal.

But shouldn't virtuous people have been willing to sell the ice cheap, hang the expense? No. Selling ice cheap merely would have resulted in a run on ice, and those

who wanted it to chill their beer might have bought it all up before those who needed it to chill their insulin even got to the head of the queue. Even someone thinking only of others' needs would find little reason to truck ice into an area "protected" by even the most lenient anti-gouging laws.

By contrast, when prices rise, buyers have a reason to conserve, sellers have a reason to supply, and the resulting competition brings prices down. Yet there is a much better third way: stockpiling generators and fuel when prices are still low, for instance. Residents of hurricane-prone areas might do this as households, cooperatives, or a public body, but in any event the point is to do something feasible. "A good society pulls together," Sandel says, "and people look out for each other."[45] True, but guaranteeing we run out of ice during an emergency is not pulling together. We look out for each other by putting feasibility first. Surely civic virtue requires that much.

How means justify ends

We come now to one final case:

> Using funds that have recently become available to us, we seek to reduce the number of deaths per year among young people. We rank causes of these deaths by the number of deaths attributable to each of them. At that point, (a) we might use our funds to focus on the leading cause of early deaths, or (b) we might focus instead on some cause farther down the list.

Here our end, generally speaking, is the rather indeterminate one of reducing the number of early deaths; options (a) and (b) offer a couple of ways in which we might make that end determinate enough that we can actually start pursuing it. It might seem obvious that option (a) gives the right way to make our end determinate; paradoxically, though, pursuing this end might actually save fewer young lives than if we had set our sights on causes farther down the list. It all depends on the available means. For instance, perhaps we do not know much about how to prevent the No. 1 cause of early deaths, while a recent medical discovery has provided a low-cost cure for a disease that is only the No. 6 cause.[46]

We are more likely to achieve the success we are after if we look at what is really feasible in the world as we find it, and *then* figure out what our determinate end should be. Look at it this way. Ends do not always justify means: reducing the No. 1 cause of early deaths is a noble end, but it may not justify the means of diverting resources away from preventing a lot more early deaths from other, more feasibly treated causes. As Steven Rhoads puts it, "We can know how high to set any one objective only if we know what we give up in progress towards other objectives."[47] Precisely because some things are precious, sometimes it has to be feasibility that tells us what is right. When it comes to making ends determinate – that is, where the rubber actually meets the road –

sometimes it really takes means to justify ends.

Conclusion

I still am not a utilitarian. I do believe that there is no virtue without practical intelligence, and that practical intelligence involves careful thinking about consequences. However, practical intelligence must also understand when consequences are not the point. Here is an analogy. Michelangelo spent four years completing his work on the ceiling of the Sistine Chapel, and one would hope that he spent none of that time worrying about the diminishing marginal value of each hour he spent perfecting his masterpiece. Michelangelo's ability to create a masterpiece like that is what made his time so valuable in the first place, so to ask whether the ceiling was worth the hours spent would have been to fail to understand what those hours were for.[48] The same is true for the virtues. When Martin Luther refused to recant his writings at the Diet of Worms, this was not because he thought the benefits of his refusal would outweigh the costs (even if he did believe they would), but because without fidelity to his conscience none of his actions would ever be worth anything. There comes a time when that is what it is like to be courageous.[49]

Costs and benefits are not always the point. It takes practical intelligence to tell the difference, because it is practical intelligence that grasps what it is to live well in an overall way. That is why I am not a utilitarian. But results *are* the point sometimes – *most* of the time, in fact – and that is why virtue ethics can learn so much from utilitarianism. One of the chief things to learn, I have argued, concerns the difficult boundary between meaning well and doing well. Meaning well, we might set our priorities and then try to make our first choice feasible. When we actually do well, though, it is usually because we first understand what is feasible and then set our priorities. A good society therefore depends on more than knowing what works and knowing what matters. It depends on knowing those things in that order.

I would like to thank Ben Eggleston and Dale Miller for the opportunity to contribute to this volume and for their generous editorial assistance, as well as Mark LeBar, David Schmidtz, and Matt Zwolinski for their comments on an earlier draft.

Notes

1. See especially Bentham, *IPML*, chapter 1, sections 6–10; Mill, *Utilitarianism*,

Collected Works, vol. x, p. 210 (chapter 2, para. 2); and Moore, *Ethics*, chapter 1. See also Sinnott-Armstrong, "Consequentialism."

2. This characterization is given by Zwolinski and Schmidtz, "Environmental Virtue Ethics."

3. Sinnott-Armstrong, "Consequentialism."

4. Consider, e.g., Sandel's complaint against cost–benefit analysis that it employs "utilitarian logic" (*Justice*, p. 41).

5. E.g. Sandel (*Justice*, p. 33 and p. 103) is not atypical in this regard.

6. I have in mind Kant's thesis in *The Metaphysics of Morals* II (a.k.a. *Doctrine of Virtue*) that we have a moral duty to make promoting the happiness of other people one of our goals. I think this idea is also connected to his claim in the *Groundwork* that the only sort of world we can rationally wish to have is one in which we develop our skills and cooperate so as to make a world we can be happy in (see R. Johnson, "Kant's Moral Philosophy," section 5, for discussion).

7. I thank Matt Zwolinski for this way of putting the point.

8. See Bentham, *IPML*, chapter 1, sections 13–14 and especially the July 1822 addition to the note on section 13.

9. Aristotle, *Nicomachean Ethics* VI.12, 1144a7–9.

10. Aristotle, *Nicomachean Ethics* IV.1.

11. Aristotle, *Nicomachean Ethics* VI.12–13. See Hursthouse, "Practical Wisdom," for an excellent discussion.

12. See Forero, "Leaving the Wild, and Rather Liking the Change."

13. For this and similar analogies, see Aristotle, *Nicomachean Ethics* III.3 and VI.12.

14. Aristotle, *Nicomachean Ethics* VI.5, 1140a25–28.

15. See Russell, *Practical Intelligence and the Virtues*, chapter 1.

16. See Hursthouse, "The Virtuous Agent's Reasons."

17. See Aristotle, *Nicomachean Ethics* VI.5, 1140a25–31 and b4–11.

18. See also Foot, *Virtues and Vices and Other Essays in Moral Philosophy*, pp. 5–7.

19. For an excellent discussion of opportunity costs, see Rhoads, *The Economist's View of the World*, chapter 2; see also Schmidtz, "A Place for Cost–Benefit Analysis," pp. 155–156.

20. The example is adapted from Slote, *Morals from Motives*, pp. 39–40.

21. See Hursthouse, "Practical Wisdom," pp. 301–302.

22. On the phenomenon of substituting easier questions for harder ones, see Kahneman, *Thinking, Fast and Slow*, chapter 9. For discussion in the context of economic reasoning, see Caplan, "Eureka! Economic Illiteracy as Mental Substitution."

23. Rhoads, *The Economist's View of the World*, pp. 16–17.

24. For a transcript of the debate, see http://archives.cnn.com/TRANSCRIPTS/1109/12/se.06.html.

25. On dilemmas in virtue ethics, see Hursthouse, "Applying Virtue Ethics" and *On Virtue Ethics*, chapters 1–3.

26. I thank David Schmidtz for this way of stating the point.

27. Cf. Hursthouse, "Practical Wisdom," pp. 290–293 and pp. 297–298.

28. See Schmidtz, "A Place for Cost–Benefit Analysis," pp. 160–161.

29. Schmidtz, "A Place for Cost–Benefit Analysis," p. 153.

30. On this point, see Roderick Long's remarks at http://bleedingheartlibertarians.com/2011/09/the-libertarian-three-step-program.

31. See G. A. Cohen, *Self-Ownership, Freedom, and Equality*, pp. 94–101.

32. Cohen sees no dilemma, arguing that Able's loss of self-direction is of little moral significance anyway (*Self-Ownership, Freedom, and Equality*, chapter 10).

33. Bradford, *Of Plymouth Plantation*, chapter 6 and entry "Anno Dom: 1623."

34. See Sandel, "How Markets Crowd Out Morals" and "What Isn't for Sale?"

35. Smith, *The Wealth of Nations*, vol. I, book I, chapter 2.

36. On the informational value of prices, see Hayek, "The Use of Knowledge in Society."

37. Aristotle, *Politics* II.5.

38. Plato, *Republic* II, 369b–373d.

39. See G. Hardin, "The Tragedy of the Commons."

40. J. S. Mill, *On Liberty*, *Collected Works*, vol. XVIII, chapter 4.

41. Sandel, "How Markets Crowd Out Morals."

42. Gaus, "On the Difficult Virtue of Minding One's Own Business," p. 24.

43. Here I am greatly indebted to Zwolinski, "The Ethics of Price-Gouging."

44. Sandel, *Justice*, pp. 7–8.

45. Sandel, *Justice*, p. 7.

46. Rhoads, *The Economist's View of the World*, p. 32.

47. Rhoads, *The Economist's View of the World*, p. 33.

48. Rhoads, *The Economist's View of the World*, discusses such cases toward the end of chapter 3.

49. I thank David Schmidtz for this point.

14 Utilitarianism and fairness

Brad Hooker

Is utilitarianism potentially in conflict with fairness? To answer this question, we need to distinguish between different kinds of utilitarianism and between different ideas about fairness. In the next three sections, I make these distinctions. In the subsequent two sections, I discuss the possibility of conflict between different kinds of fairness and different kinds of utilitarianism.

Different kinds of utilitarianism

When early utilitarians referred to "the greatest good of the greatest number," they might have had in mind not only the maximization of aggregate welfare but also the spreading of benefits to as many individuals as possible. In cases where what would produce the greatest aggregate welfare would not be what would benefit as many individuals as possible, which is more important? Whatever answer early utilitarians would have given, at least since 1960 the term 'utilitarianism' has usually referred to a moral and political philosophy that evaluates either acts or rules or both purely in terms of their effects on aggregate welfare. Utilitarian calculation of aggregate welfare is impartial in that benefits or harms to any one individual are counted exactly the same as benefits or harms of the same size to any other individual.

The kind of utilitarianism most often discussed in textbooks is the kind that conflicts most sharply with non-utilitarian theories. This kind of utilitarianism is maximizing act utilitarianism. According to maximizing act utilitarianism, (1) an act is morally *required* if, only if, and because it would produce greater aggregate welfare than any alternative act, (2) an act is morally *permissible* if, only if, and because no alternative act would produce greater aggregate welfare, and (3) an act is morally *wrong* if, only if, and because at least one other alternative act would produce greater aggregate welfare.

In the rest of this chapter, I will regularly refer just to a theory's account of what is morally required, or of what is morally permitted, or of what is morally wrong, but not all three. This is just to save words. Each theory I will discuss will have an account of all three, unless I indicate otherwise.

Another kind of act utilitarianism is satisficing act utilitarianism, which is the view that an act is morally permissible if, only if, and because the act produces enough aggregate welfare, the presumption being that enough aggregate welfare is something short of the maximum available.[1] Satisficing act utilitarianism is like maximizing act utilitarianism in holding on to the distinctions between the morally required, the morally permissible, and the morally wrong. But satisficing act utilitarianism draws the distinctions in different

places than maximizing act utilitarianism does. For maximizing act utilitarianism, the line between the permissible and the wrong is drawn right at the top of the scale of good outcomes – only acts which produce at least as much utility as any other act are morally permissible. For satisficing act utilitarianism, in order to be permissible, acts have to produce enough utility but do not have to produce the maximum available.

One of the great attractions of satisficing act utilitarianism in comparison with maximizing act utilitarianism is that satisficing act utilitarianism leaves the agent with a larger range of morally permissible alternatives. In every situation, maximizing act utilitarianism shrinks the set of morally permissible actions down to the single act that maximizes welfare, unless there are two or more equally maximal acts in which case maximizing act utilitarianism requires the agent to choose one of these. Shrinking the set of morally permissible actions so far is implausibly restrictive, leaving the agent with vanishingly little moral freedom to choose between morally permissible acts.[2] In contrast, depending on where satisficing act utilitarianism draws the threshold of "good enough," there will be many more than one or two possible acts that pass this threshold and thus qualify as morally permissible according to satisficing act utilitarianism.

The other great attraction of satisficing act utilitarianism is that it typically requires less self-sacrifice of the agent than does maximizing act utilitarianism. Maximizing act utilitarianism requires the agent to keep sacrificing for the sake of others up to the point where further sacrifices will harm the agent more than they will benefit others. Satisficing act utilitarianism requires merely that the agent make whatever sacrifices are needed to produce enough aggregate welfare, but this will typically be somewhat less than what would be required to maximize aggregate welfare.

While these two attractions of satisficing act utilitarianism are very powerful, there are some devastating objections to satisficing act utilitarianism. The most obvious difficulty with satisficing act utilitarianism is about how it can draw a *non-arbitrary* line between enough aggregate welfare and not enough aggregate welfare. Even more compelling is Tim Mulgan's objection that satisficing act utilitarianism permits the agent to choose what benefits others less than the maximum possible even when choosing what would benefit them the most would be no more costly or difficult for the agent.[3]

Another kind of act utilitarianism is scalar act utilitarianism.[4] This view jettisons the distinctions between the morally required, the morally permissible, and the morally wrong in favor of an undifferentiated scale from best possible act to worst possible act. Thus, according to scalar act utilitarianism, no acts are morally required, morally permissible, or morally wrong. Scalar act utilitarianism holds instead that there is the act that would produce the greatest aggregate welfare, and there is the act that would produce the second greatest aggregate welfare, and there is the act that would produce the third greatest aggregate welfare, and so on down the scale to the act that would result in the least aggregate welfare.

The advantage scalar act utilitarianism is supposed to have over maximizing act utilitarianism is that scalar act utilitarianism does not have the burden of defending the extremely restrictive and demanding standard of morally permissible action that maximizing act utilitarianism avows. One advantage scalar act utilitarianism has over satisficing act utilitarianism is that scalar act utilitarianism does not have the burden of specifying what counts as *enough* aggregate welfare. Another advantage of scalar act utilitarianism over satisficing act utilitarianism is that scalar act utilitarianism does not have to contend that an agent's knowingly choosing a good but suboptimal outcome for others is morally permissible even when choosing the optimal outcome for them would involve no greater sacrifices from the agent.

But scalar act utilitarianism can take such positions only because it has abandoned altogether the concepts of moral permissibility and moral wrongness and thus the distinction between morally permissible acts and morally wrong acts. This distinction is absolutely central to moral thought. Much moral thought is focused on what moral duty requires of us, as opposed to what is permitted but not required. And many moral reactive attitudes pivot on the distinction between wrongness and permissibility. For example, an act cannot be morally blameworthy unless it was morally wrong; and feelings of guilt, resentment, or indignation are out of place if what was done was morally permissible.

From now on, the only variety of act utilitarianism I will discuss is maximizing act utilitarianism. My reason for this is that maximizing act utilitarianism has been far more prominent than both satisficing act utilitarianism and scalar act utilitarianism.[5]

Act utilitarianism can be formulated in terms of actual results or in terms of probabilities. Act utilitarianism formulated in terms of actual results holds that these results determine what was morally required, permissible, or wrong. Act utilitarianism formulated in terms of probabilities starts by listing the possible benefits and possible harms of an act. Then it multiplies the utilities of those possible benefits and harms times the probabilities of their occurring if the act is done. Then it sums these numbers to reach the "expected utility" of the act.

Here is a standard example. Suppose a doctor could give her patient a pill that has a 99.9-percent chance of curing what would otherwise be a fatal cancer and a 0.1-percent chance of killing the patient immediately. Suppose the doctor then gives the patient the pill and the patient dies. The actual utility of giving the patient the pill was very low, indeed very negative. But the expected utility of giving the patient the pill was very high.

Of course we would regret the doctor's giving this patient this pill. But we certainly would not blame the doctor for doing what she sincerely and reasonably believed had a 99.9-percent chance of curing the patient. We might even say that it would be *unfair* to blame the doctor for being so terribly unlucky. In fact, we blame people for taking terrible risks with other people's welfare, even when luck prevailed and the probable

harms did not in fact occur. And we would blame the doctor if she did not prescribe the medicine even if we were somehow able to determine that it would have killed the patient instantly, as long as the doctor had no reason to suspect that it would have that result.

Because of the additional potential for unfairness if act utilitarianism is formulated in terms of actual utility, such thoughts militate in favor of casting any discussion of relations between utilitarianism and fairness in terms of expected utility instead of actual utility. This is what I shall do.

Unlike all forms of act utilitarianism, rule utilitarianism assesses acts in terms of rules. According to rule utilitarianism, the rules with the greatest expected utility are justified, and acts are justified if and because they are allowed by these rules.

If poorly formulated, rule utilitarianism runs into fatal objections. On some formulations, rule utilitarianism does not even manage to disagree with maximizing act utilitarianism about which actions are morally right. However, this chapter is not about how to formulate rule utilitarianism, nor about the multitude of possible objections to rule utilitarianism.[6] In the present chapter, without explaining why, I will simply presuppose that the best formulation of rule utilitarianism is that an act is morally permissible if, only if, and because the act is allowed by the code of rules whose internalization by the overwhelming majority of everyone in future generations would maximize expected aggregate welfare. And, to save words, I will often use the term "optimific rules" to refer to rules whose internalization by the overwhelming majority of everyone in future generations would maximize expected aggregate welfare.

Now if utilitarianism is understood as a family of theories each of which evaluates either acts or rules or both purely in terms of the aggregate welfare, how should we classify the member of this family that evaluates *both acts and rules* purely in terms of aggregate welfare? In some circles, this theory is called *global utilitarianism*.[7]

Suppose *the rule* whose acceptance would maximize aggregate welfare could, in at least some situations, require *an act* that would not maximize aggregate welfare. Global utilitarianism would endorse accepting this rule, since by hypothesis its acceptance is welfare maximizing. At the very same time, global utilitarianism would condemn acts of compliance with the rule in these situations, since by hypothesis those acts are not welfare maximizing.

Act and rule utilitarians disagree primarily about what makes *acts* right or wrong. On this matter, global utilitarians agree completely with act utilitarians. Thus, global utilitarianism is a kind of act utilitarianism.

Many prominent act utilitarians, e.g., Henry Sidgwick and J. J. C. Smart, have been global utilitarians.[8] Could one be an act utilitarian without being a global utilitarian? Yes, one could be an act utilitarian about acts and yet assess rules not in terms of the utility of their acceptance but in terms of the frequency with which these rules would result in

right (utility-maximizing) acts. But such possibilities need not bother us here. For the purposes of this chapter, what matters about global utilitarianism is its act-utilitarian assessment of acts. There is thus no need to mention global utilitarianism again in this chapter.[9]

Different kinds of fairness

Just as we need to distinguish between different kinds of utilitarianism, we need to distinguish between different kinds of fairness.

Rules and principles make distinctions. Formal fairness consists in the equal and impartial application of these distinctions. For example, if there is a rule that people who earn over $100,000 have to pay 40 percent in tax, then formal fairness is violated if the senator's husband is allowed to pay a lower rate of tax though he earns over $100,000. If there is a rule that the first one in line gets served first, then formal fairness is violated if, though the first in line is normally served first, this isn't true when the first in line has an unpopular accent.

Formal fairness – the equal and impartial application of the same rules to everyone – is not exhaustive of fairness; rules can be unfair in their content. Whether or not rules are applied equally and impartially, they can make distinctions that fairness forbids. To illustrate, imagine a set of rules distributing advantages or disadvantages purely on the basis of eye color or gender. The equal and impartial application of such rules would be *formally fair* but *substantively unfair*. In order for rules to have fair content, the distinctions between people made by these rules must be ones that fairness allows to be made or even demands to be made. Distinguishing between people on the basis of eye color or gender is not allowed by fairness.

It is much easier to show that we must distinguish between formal and substantive fairness than to say what determines substantive fairness. There are a number of rival views. I catalogue the most prominent in the following section.

Different views of substantive fairness

A social practice is partly constituted by rules, which then frame expectations, intentions, reactions to noncompliance, etc.[10] Social practices have many different purposes, of course. Nevertheless, a common idea is that a social practice's rules are fair depending on whether they make distinctions that serve the point of the social practice.

Formulated broadly as a claim about *every* social practice's rules, that idea is implausible. For some social practices have indefensible purposes. A rule could not be rendered fair simply by the fact that it makes distinctions that serve the point of an indefensible social practice. Imagine a social practice the purpose of which is to make

some people feel superior to others or inferior to others on the basis of the social standing of their parents. The rules of this social practice might dictate that anyone whose parents did not have high social standing should bow and defer to anyone whose parents did have high social standing. This rule would serve the purpose of the social practice in question, but would not be making a distinction that fairness allows.

So we cannot accept the broad idea that *any* social practice's rules are fair depending on whether they make distinctions that serve the point of the social practice. We should thus consider a narrower idea focused on social practices with defensible and thus permissible purposes. This is the idea that any distinctions that serve permissible purposes of social practices are permitted by fairness.

I do not mean to suggest that social practices have permissible purposes only when those purposes serve fairness. Many social practices have the purpose of efficiently producing goods that are specifiable independently of the social practice and do not include fairness. Such goods might be knowledge or health or enjoyment. Many social practices have the purpose of paying homage to independently specifiable goods. An example is the social practice of congratulating teachers on the success of their former students. And many social practices have the purpose of specifying a kind of excellence and the conditions under which this excellence can be achieved. Athletic and intellectual games and artistic genres of many sorts provide examples.

Since social practices are permitted to have so many different purposes and many of these overlap or complement one another, consider the idea that society is a network of social practices. Must society achieve certain things in order for its social practices (and their constituent rules) to be fair? In the remainder of this section, I will outline a number of the possible alternative answers to this question. In later sections, I will say how conceptions of fairness can conflict with act utilitarianism or with rule utilitarianism.

One idea is that the social practices of a society are fair if and only if *the social practices have maximum expected utility*. The line of thought might be that utilitarian assessment of social practices and rules has, at the foundational level, equal concern for the welfare of each. Since the foundation seems so eminently impartial, is not this foundation fair? If this is a fair foundation, then would not the social practices and the acts those social practices license inherit fairness from the foundation?

A second and very different idea is that the social practices of a society are fair if and only if the social practices require each to contribute equally to public goods as long as others are contributing. Whereas the previous idea suggested that fair practices are whichever ones maximize expected utility, the second idea ties fairness to *reciprocity*. The complaint that existing social arrangements allow "free loaders" and "tax cheats" to piggyback on the hard work of others testifies to the underlying appeal of this conception of fairness.

A third and again very different idea is that the social practices of a society are fair if and only if they are geared to make sure that people get what they *need*, at least where

the society has enough resources to provide this. It is often said to be manifestly unfair that some have more than they need while others have less than they need.

A fourth idea is that social practices are fair if and only if they *leave the worst off as well-off as the worst off under any other arrangements would be left*. This idea was made popular by John Rawls.[11] His "difference principle" held that inequalities are fair only if they leave the worst off better off than the worst off without such inequalities would be left. Rawls's difference principle implies the requirement to "*maximin*" (for "maximize the minimum position").

Maximin was justified, Rawls argued, by the manifest fairness of a rational choice situation that Rawls took over from John Harsanyi and then modified. Suppose that, instead of trying to select rules from a point of view in which there is equal concern for everyone and full information about the likely benefits and harms of different possible rules, we select rules for society from an "original position" in which we care about only ourselves but are behind a "veil of ignorance" which hides from us all specific information about ourselves. Behind this veil, we do not have any idea whether we are talented, energetic, healthy, female, religious, etc. Since behind the veil we have no information that could bias our selection of rules, Rawls proposes the rules we select behind the veil would be selected fairly and impartially.

Harsanyi argued that the rational choice behind the veil of ignorance would be to choose whatever rules would maximize utility. If one had an equal chance of being anyone once the veil went up, then the way to maximize one's own expected utility behind the veil would be to pick the rules with the greatest expected average utility, everyone's utility being counted equally and impartially. Rawls argued against Harsanyi, however, that behind the veil of ignorance one would be rational to be risk averse and thus focus on the position of the worst off instead of the average position.

That argument of Rawls's has widely been thought to be unpersuasive. Why exactly does rational choice require risk aversion? So the argument from the veil of ignorance to maximin is highly contentious.

Just as contentious is whether maximin is too strong to be plausible. Suppose you face a choice between directing very small benefits to the worst off (e.g., an additional few *hours* of healthy life) or directing very large benefits to those a little better off already (e.g., ten additional *years* of healthy life). Suppose that these are your only options. Here, it seems hard to believe that fairness insists on the comparatively tiny benefits for the worst off over very large benefits for the better off.

In the face of such objections, we might try taking the hard edge out of maximin. A view called *weighted prioritarianism* gives some degree of priority to benefits to the worse off over benefits to the better off, without going so far as to give overriding priority to benefits to the worse off.[12] How much more does a benefit to the worse off count than a benefit of the same size to the better off? Answers are varied and vague.

Nevertheless, the idea does seem deeply attractive.

To see what is attractive about it, suppose we are evaluating two possible distributions. In the first of these distributions, there is great aggregate utility but also huge inequality between the best off and the worst off. In the other distribution, there is a bit less aggregate utility but the worst off are much better off than the worst off are in the first distribution and the inequality between the worst off and the best off is much less.

What is the best possible explanation of the fact that the second distribution is morally better? This explanation cannot be that the second distribution has greater utility, because it does not. The best explanation cannot be merely that equality of welfare is higher in the second, since equality of welfare is not always desirable (e.g., where equality of welfare can be achieved only by leveling down the welfare of the best off to that of the worst off). The best explanation cannot be supplied by maximin, since that principle is not reliably right in its implications (as we have just seen). The best possible explanation of why a distribution with lower aggregate utility but greater benefits for the worst off is morally better is that, from an impartial point of view, benefits for the worse off matter more than benefits for the better off that are smaller, the same size, or even a bit bigger than the benefits for the worse off.

So far, I have outlined five different ideas about what makes rules and social practices substantively fair: (1) the idea that rules and social practices are fair when they maximize expected aggregate welfare calculated in such a way that each counts equally, (2) the idea that rules and social practices are fair when they require reciprocity, (3) the idea that rules and social practices are fair when they ensure that everyone gets what they need insofar as this is possible, (4) the idea that rules and social practices are fair when they maximin, i.e., leave the worst off as well-off as they would be under the best of the alternative rules and practices, (5) the idea that rules and social practices are fair when they maximize expected utility where this is calculated in a weighted prioritarian way, i.e., by giving some degree of extra weight to benefits for the worst off. A sixth conception is (6) the idea that the social practices of a society are fair if and only if people get what they deserve.

Conflicts between utilitarianism and formal fairness

Formal fairness and act utilitarianism can conflict. Suppose a rule is promulgated that medicines will be distributed to those who need them on a first-come, first-served basis. The first ill person is given medicine. The second ill person is given medicine. The third ill person, however, is not given medicine, because he will in fact waste it, to his own detriment. So we do not here have formal fairness concerning the rule that medicines will be distributed to those who need them. And, if the medicine that could have been given to the third person were instead given to some other ill person who would benefit from it, this infringement of formal fairness concerning the rule that medicines will be distributed

to those who need them would in fact maximize welfare.

One reaction to this example might be to say that of course the equal and impartial application of any rule other than "choose whatever act maximizes expected utility" can get in the way of maximizing utility. The problem might not really be a conflict between formal fairness and act utilitarianism but rather a conflict between act utilitarianism and any alternative to act utilitarianism. After all, there is no necessity of conflict between formal fairness and the rule "choose whatever act maximizes expected utility."

That reaction is too glib. Even act utilitarians admit that constantly making decisions by trying to apply their act-utilitarian principle will not maximize utility. People often lack information about the probable effects of their choices and, without such information, could not calculate expected utility. Furthermore, obtaining such information would often involve greater costs than are at stake in the decision to be made. And, even if people trying to make decisions had the relevant information about possible benefits and harms and their probabilities, calculating expected utilities is typically unpleasant and time-consuming and thus a cost in itself. And the agent might make mistakes in the calculations. (This is especially likely when the agent's natural biases intrude, or when the calculations are complex, or when they have to be made in a hurry.)

Finally, there are what we might call expectation effects. Imagine a society in which people know that others are naturally biased toward themselves and toward their loved ones but are trying to make their every moral decision by calculating aggregate utility. In such a society, each person might well fear that others will go around breaking promises, stealing, lying, and even assaulting whenever they convinced themselves that such acts would maximize utility. In such a society, people would not feel they could trust one another.

For all of the reasons above, even philosophers who espouse the act-utilitarian criterion of moral wrongness reject an act-utilitarian decision procedure. In its place, they typically advocate that, at least normally, agents should decide what to do by applying rules such as "Do not harm innocent others," "Do not steal or vandalize others' property," "Do not break your promises," "Do not lie," "Pay special attention to the welfare of your family and friends," and "Do good for others generally." Hence, formal fairness in the application of the rules that act utilitarianism tells us to use as our regular decision procedure can conflict with act utilitarianism.

Rule utilitarianism holds that the connection between rules and moral rightness is much tighter than act utilitarianism holds that it is. Act utilitarianism accepts that agents should run their daily lives by applying the familiar rules just mentioned, but act utilitarianism holds that these rules play no part in determining what is morally required, morally permissible, or morally wrong. So act utilitarianism certainly does hold that which acts are morally required are often not the ones that would be selected by the equal and impartial application of the rules just mentioned. Rule utilitarianism, in contrast, holds that what is morally required, morally permissible, or morally wrong is

determined by whatever rules are the ones whose acceptance would maximize expected utility. Because of the much more central role that rule utilitarianism accords to rules in determining moral verdicts, rule utilitarianism will conflict far less than act utilitarianism with the equal and impartial application of rules such as "Do not harm innocent others," "Do not steal or vandalize others' property," "Do not break your promises," "Do not lie," "Pay special attention to the welfare of your family and friends," and "Do good for others generally."

Conflicts between utilitarianism and substantive fairness

As I indicated, there are a variety of different ideas about what grounds substantive fairness.

The first one I listed was that the social practices of a society are fair if and only if the social practices have the greatest expected utility. The social practices with the greatest expected utility might sometimes require actions with lower expected utility than alternatives. This is why act utilitarianism and rule utilitarianism do not always agree about which acts are required, which are permissible, and which are wrong. For, in such cases, act utilitarianism will of course require *the act* with the highest expected utility, while rule utilitarianism will require conformity with *the social practice* that has the highest expected utility. Now, if substantive fairness is determined by the social practices that utilitarianism favors, then, since rule utilitarianism also requires conformity with these practices, there is no conflict between rule utilitarianism and fairness, even when these social practices require actions with lower expected utility than alternative acts. But in these cases there is conflict between act utilitarianism and fairness.

The second conception of substantive fairness I listed was the view that the social practices of a society are fair if and only if the social practices require each to contribute equally to public goods as long as others are contributing. David Lyons argued that this conception of substantive fairness conflicts with both act utilitarianism and rule utilitarianism, though under different conditions.[13]

Suppose enough others are paying their taxes, and I can get away with not paying mine without threat to the public good and without getting punished. My own welfare will be greater if I do not pay taxes, and my paying my taxes would not increase aggregate welfare. In such a case, act utilitarianism would apparently tell me to keep the money for myself rather than pay my taxes. But fairness definitely requires that, since others are paying their taxes, I should pay mine too. So the objection is that act utilitarianism cannot explain what is wrong with one person's free riding on the good behavior of others when this person's free riding costs no one anything.

There is some controversy about whether this argument is correct about what act

utilitarianism requires. Derek Parfit, for example, argued that, if we avoid various common mistakes in moral mathematics, we will reach the conclusion that the way to maximize expected utility is not to free ride on the contributions of others but instead to do our part.[14] But whether Parfit's arguments here are sound has been controversial.[15] I think these arguments of his have also received less attention than most of his other work and that the reason for this is that the objection that act utilitarianism can prescribe not contributing to a public good when there are enough contributions already is not as compelling as many of the other objections to act utilitarianism, such as the objection that act utilitarianism requires one to harm innocent others when this is necessary to maximize aggregate utility and the objection that act utilitarianism's requirement about making sacrifices for others is implausibly demanding.

Unlike act utilitarianism, rule utilitarianism can explain why I should comply with optimific rules, even when my compliance will be costly to myself and not beneficial to anyone else. As Richard Brandt wrote,

> there could hardly be a public rule permitting people to shirk while a sufficient number of others work . . . It would be all too easy for most people to believe that a sufficient number of others were working . . . Would it even be a good idea to have a rule to the effect that if one absolutely knows that enough others are working, one may shirk? This seems highly doubtful.[16]

The problem cases for rule utilitarianism are ones where most others are not following the optimific rules.[17] In these cases too, my complying with the optimific rules would be costly to me. So again act utilitarianism tells me not to follow them, unless the benefits to others would outweigh the costs to me. But, while in the earlier cases where others *are* complying with the optimific rules it seems morally required for me to comply with them too, in the present cases where others are *not* complying with the optimific rules it seems morally permissible for me not to comply with them either. So the objection is that rule utilitarianism tells me to comply with optimific rules, whether or not others are in fact complying with those rules, that is, whether or not others are reciprocating my compliance.

In cases where my complying with a rule would benefit the very people who are not themselves complying with it, then an insistence that I nevertheless comply would threaten to turn me into a "sucker" ripe for exploitation by the "cheats."[18] Such a requirement on compliers in a world of noncompliance would undeniably be unfair on the compliers and advantageous to the noncompliers.

It would be counterproductive to require people to comply with rules where this compliance is beneficial to those who refuse to follow these rules. One of the best ways of discouraging other people's noncompliance is to make one's complying with such rules contingent on their reciprocal compliance. Rule utilitarianism would thus advocate a rule

allowing one to deter the noncompliance of others by refusing to comply unless they do so as well.

But what about cases where noncompliers would not be the beneficiaries of one's compliance with rules? Cases in which people can make contributions to responsible rescue agencies are good examples. Suppose I believe that if all accepted a rule requiring the comfortably off to give at least 2 percent of their income to responsible rescue agencies, expected utility would be maximized. But suppose I know that most other comfortably off people will not give anything like this much. If I nevertheless give at least 2 percent of my income, the benefit will go to those needing rescue, not to the comfortably off people who are not doing their share. Requiring that I give at least 2 percent of my income is not unfair on me and does not unfairly benefit those who are in a position to comply with this rule but do not.

I have argued elsewhere that rule utilitarianism would not tell agents to ignore whether others are doing as they ought. Rule utilitarianism would instead tell agents to be aware of whether others are doing as they ought – and if they are not, then to be willing to make additional sacrifices to protect potential victims from others' noncompliance.[19] The benefits of having people be willing to do extra to make up to some extent for the noncompliance of others are obvious. More people will be rescued than would be if people were not willing to make up to some extent for the noncompliance of others.

It might seem unfair on those who make their fair-share contributions to be required to contribute more than their fair share to make up for others who equally could contribute but are not doing so. This is the line of thought developed in Liam Murphy's *Moral Demands in Nonideal Theory*.[20] Perhaps rule utilitarians can provide a rationale for a rule that initially requires each who is equally able to contribute to contribute an equal amount, which is the person's fair-share contribution, and then requires people to make up to at least some extent for the noncompliance of others when that occurs. The rationale comes from the diminishing marginal utility of material goods and other resources, as I will now explain.

Normally, the more food or clothes or money or energy people have, the less they benefit from gaining an additional unit. There are many exceptions to this generality. For example, that additional unit of money might put you over some important threshold, such as enabling you to buy something you need. Or that additional bit of energy might enable you to surge past the other competitors and win the race. But, as a generality, it is true that material goods and other resources have diminishing marginal utility.

Now if you and I are equally comfortably off, then there is a presumptive case for utility's being maximized by your and my giving equal shares to responsible rescue agencies. Suppose I give nothing rather than my equal share, and so you give an additional amount to make up for my failure. Other things being equal, the utility loss to you of the additional contribution will be more than the loss to me would have been of my giving my equal share.

On the other hand, as I indicated, there are obvious benefits of having people be willing to make additional contributions to rescue agencies to make up to at least some extent for the noncompliance of others. So there are utilitarian reasons to favor a rule requiring people to make up to at least some extent for the noncompliance of others in such cases. There may also be reasons of fairness to require contributions beyond one's fair share when others are not giving their fair share, as I will now explain.

Perhaps even more fundamental than the concept of fair-share contributions to collective projects (such as funding rescue agencies) is the idea that fairness calls for society to be configured in ways that ensure that, to whatever extent possible, everyone gets what he or she really needs. The idea is that it is unfair for some people to have far more than they need while others do not have enough to meet their needs. So if I have more than I need and others have not enough to meet their essential needs, then perhaps there are reasons of fairness to require redistributions from me to those people. These reasons of fairness have to do with my relations to those in need, not with my relations to others who are not in need but instead are in just as good a position as I am to help the needy. So there are reasons of fairness for me to make contributions to help the needy even if others are not giving their fair share.

The idea that fairness requires the satisfaction of needs insofar as possible, if this idea is to have even initial plausibility, must rely on a distinction between what is necessary to avoid death, disability, social exclusion, etc., and what is instrumental to the fulfillment of merely optional preferences. You need about 400,000 calories over the next year to avoid starving to death. You do not actually need a straighter nose, whiter teeth, or a more luxurious car, though of course you might prefer all these things. If there are enough resources in society for everyone to have what they need to avoid death, disability, social exclusion, etc., then perhaps fairness requires that you get the calories you need to survive. But fairness certainly does not insist that you be given the resources necessary to get your nose straightened, your teeth whitened, and your car upgraded.

Obviously, an account is needed of the distinction between *merely optional* preferences and *non-optional* needs. A familiar way of drawing this distinction is to say that your merely optional preferences do *not* have to be fulfilled in order for you to avoid harm but your needs *are* things that must be fulfilled in order for you to avoid harm.[21]

Here 'avoid harm' must mean something other than "be worse off than you were before." Suppose there is some good you now lack but used to have. You were better off before you lost this good. All this can be true without entailing that you *need* this good or that you are *harmed* by not having it. You might have been better off, for example, before your face started getting wrinkled. But it would be hyperbolic to say that you *needed* a miracle prevention of wrinkles before you started getting them or that you were *harmed* by going from smooth to wrinkled.[22]

Likewise, for the purposes of defining the relevant sense of 'need', 'avoid harm'

cannot be defined as "be worse off than you would have been otherwise." The reason is that, because you lacked X, you might be worse off than you would have been otherwise and yet you might not have really needed X and are not harmed by lacking X. For example, you might be worse off now than you would have been had salaries not been frozen, and yet the freezing of salaries was not something that you actually needed not to happen or something that harmed you.

When 'need' is defined as that which someone must have in order to avoid harm, I think 'harm' needs to be defined as a lack of some essential ingredient of a decent life.[23] For example, a decent level of health is necessary in order to avoid harm and to have a decent life. In trying to give other examples, we quickly get into controversy. We might try contending that you are harmed if you lack access to social engagement, for instance. But what if you do not want social engagement or even the possibility of social engagement? How can you be harmed by the absence of something you really do not want anyway?

As popular as the view is that fairness requires that people get what they need insofar as this is possible, the concept of need is too contested and vague, mainly because where the threshold is of a decent life is too contested and vague.[24] But if we are to try to explain fairness *without* any reference to needs, is there a related concept we should employ?

As mentioned earlier, weighted prioritarianism is a more promising idea than the idea that we should always feel under some moral pressure to equalize welfare or resources, since sometimes the only way to equalize is to level down. Also as mentioned earlier, weighted prioritarianism is more intuitively appealing than maximin, since maximin would sacrifice huge benefits for the better off for the tiniest gain to those worse off. So we should now ask whether there is indeed a tight connection between weighted prioritarianism and fairness. Does fairness demand that an assessment of benefits and harms accord more importance to benefits and harms to the worse off than it does to benefits and harms of the same size (or even a bit larger) to the better off?

The diminishing marginal utility of material goods and other resources (such as energy) militates in favor of redistribution from those who have more material goods and other resources to those who have less. This is a point widely emphasized by utilitarians. But weighted prioritarianism is a view about how gains and losses to utility itself should be assessed. Weighted prioritarianism is not a view merely about how the means to utility (material goods and other resources) get used in utility's production. So the utilitarians' point that material goods and other resources typically have diminishing marginal utility must not be conflated with the contention that weighted prioritarians make that gains to the utility of the worse off matter more than smaller, the same size, or even a bit bigger gains to the better off.

There is one very powerful argument in favor of weighted prioritarianism.[25] As I

suggested above, weighted prioritarianism seems to provide the best explanation of why a somewhat smaller but more equal distribution of utility is morally superior to a somewhat larger but less equal distribution. If asked why it is morally superior, we might well say the somewhat smaller but more equal distribution of utility is fairer. And we might then say that what makes the more equal distribution fairer is that it is the one weighted prioritarianism would favor.

Arguments back and forth between utilitarianism and weighted prioritarianism (and between them and maximin) presume that there is no claim on benefits prior to, or weightier than, the size of the benefits, or how many people would get them, or the welfare levels of the parties relative to one another. But often there is a claim prior to and weightier than such considerations, namely the claim that one person *deserves* some good more than others do. If Jill deserves money or honor or privilege more than Jack does, then she ought to get it, even if he would benefit more from getting it, or even if he is worse off than she is. Recognizing the importance of this sort of consideration, very many people think that the social practices of a society are fair if and only if people get what they deserve.

There are many questions to be raised about desert. How can you deserve what you use your abilities to achieve when you did not deserve the genetic or environmental conditions which fostered the development of those abilities? Is effort the basis of desert or is productivity or is something else? Is one person's desert always only *comparative* to other people's desert, or is there *noncomparative* desert?[26]

Another important question about desert is whether all principles of desert are "institutional" or whether some are "pre-institutional." John Rawls argued that the institutional account of desert understands all principles of desert as entirely derivative from *just* institutions.[27] On this view, just institutions must be determined by criteria other than whether they result in people's getting what they deserve. The pre-institutional account of desert, in contrast, holds that, while many principles of desert make no sense without their being embedded in an institutional context or social practice, some principles of desert do *not* depend on the selection of institutions and social practices on criteria other than whether they give people what they deserve.

On the debate between institutional and pre-institutional accounts of desert, utilitarians agree with Rawls. Utilitarians evaluate actual and possible institutions and social practices purely by how much utility would result from their establishment, not by whether they would give people what they deserve. And those who advocate a maximin criterion evaluate such institutions and social practices by the results for the worst off. Weighted prioritarians assess such institutions and social practices in terms of the aggregate welfare where this is calculated with extra weight given to benefits for the worse off. All these views are versions of institutionalism about desert.

The question of whether all principles of desert are determined by institutions and social practices or instead some principles of desert are pre-institutional seems to me

more important than other questions about desert because institutionalism gives us ways of dealing with those other questions. For example, institutionalism enables us to decide whether the basis of desert should be effort, or productivity, or some combination of them, or some alternative to them. Utilitarian institutionalists, for instance, would decide by asking which reward structure would maximize aggregate welfare calculated by giving the same weight to a benefit to any one individual as to the same-size benefit to any other individual. Weighted prioritarians would decide by asking which reward structure would maximize aggregate welfare calculated by giving somewhat more weight to benefits to the worse off.

Even if utilitarian institutionalists are correct about which institutions and social practices determine the principles of desert, there is room for divergence in particular cases between what rule utilitarianism requires overall and what people deserve. For rules about desert constitute not the whole but only a subset of the rules that rule utilitarianism endorses, and, in particular cases, conflicts between rules about desert and other kinds of rules can arise. Perhaps rules about desert are particularly weighty and so normally trump whatever other rules come into conflict with them. Still, it is hard to believe that rule utilitarians will think that rules about desert always trump other kinds.

The scope for conflict between desert and utility is far greater with act utilitarianism. An act utilitarian has to believe that, even if certain institutions, social practices, and their constituent rules are perfectly well justified on utilitarian grounds, cases regularly arise in which complying with the perfectly well-justified rule about desert will predictably not maximize utility. This is true, for example, when unpopular people deserve rewards, or when people who deserve the rewards will not actually benefit much from them.

Conclusion

This chapter has distinguished between the main kinds of utilitarianism – act utilitarianism and rule utilitarianism. It has also distinguished between different kinds of fairness.

Formal fairness consists in the equal and impartial application of rules. With respect to the question of which acts are morally required, act utilitarianism regularly offers a different verdict from the one implied by application of rules other than the rule "always choose the act that will maximize utility." In this way, the equal and impartial application of rules other than "always choose the act that will maximize utility" will conflict with act utilitarianism.

Substantive fairness is a matter of the content of the rules, of whether the rules make all and only the distinctions that fairness requires. Fairness requires that a distinction be made between those who are reciprocating or are willing to reciprocate kindness toward them and those who are not willing to do this. Is the application of this distinction one that rule utilitarianism can endorse? I argued that it is. But the application of this distinction conflicts with act utilitarianism in many cases.

Fairness seems to consist of more than merely formal fairness plus reciprocity. One popular idea is that it is unfair for society to be arranged in such a way that some people end up with more than they need while others have less than they need. Where is the threshold between what is needed and what is good but not needed? Despairing of a persuasive answer to that question, I offered weighted prioritarianism as an alternative to thinking in terms of needs. Weighted prioritarianism is a rival to utilitarianism, and some objections to utilitarianism on the grounds of fairness are not also objections to weighted prioritarianism.

The concept of desert is perhaps even more central than the concept of need to popular thinking about fairness. There are many questions about the concept of desert. I suggested that the most important of these is whether all the principles of desert are derivative from institutions and social practices that are justified on grounds that do not include principles of desert, or instead whether there are some principles of desert that are "pre-institutional." If all the principles of desert are derivative from institutions and social practices that are justified on grounds that do not include principles of desert, then thinking about which possible institutions and rules would be best on those grounds might help us with those other questions about desert. However, even utilitarian institutionalists about desert will admit that there is scope for conflict between desert and utilitarianism. Again, possible conflict between rule utilitarianism and desert will be less than between act utilitarianism and desert.

I am grateful to the editors and Charlotte Newey for their comments on earlier versions of this chapter.

Notes

1. Satisficing act utilitarianism is drawn from Michael Slote. See his "Satisficing Consequentialism," part I; *Common-Sense Morality and Consequentialism*; *Beyond Optimizing*; and *From Morality to Virtue*.

2. Vallentyne, "Against Maximizing Act Consequentialism," pp. 23–28.

3. Mulgan, *The Demands of Consequentialism*, pp. 129–142.

4. See Slote, *Common-Sense Morality and Consequentialism*, chapter 5; and Norcross,

"The Scalar Approach to Utilitarianism."

5. Maximizing act utilitarianism is the focus of Chapter 6 of this volume.

6. For fairly recent discussions, see my *Ideal Code, Real World*; and Mulgan, *Future People*. See also Chapter 7 of this volume.

7. Pettit and Smith, "Global Consequentialism," and Kagan, "Evaluative Focal Points."

8. Sidgwick, *The Methods of Ethics*; and J. J. C. Smart, "Outline of a System of Utilitarian Ethics."

9. Julia Driver discusses global utilitarianism in Chapter 8 of this volume.

10. See Rawls, "Justice as Fairness," pp. 164–169 and p. 164, n. 2.

11. Most influential was his *A Theory of Justice*.

12. See Raz, *The Morality of Freedom*, p. 227; Nagel, *Equality and Partiality*, chapter 7; Parfit, "Equality and Priority"; D. Miller, *Principles of Social Justice*, pp. 223–225; and Fleurbaey, Tungodden, and Vallentyne, "On the Possibility of Nonaggregative Priority for the Worst Off."

13. See Lyons's *Forms and Limits of Utilitarianism*, pp. 139–141.

14. Parfit, *Reasons and Persons*, ch. 3.

15. See Griffin, *Well-Being: Its Meaning, Measurement, and Moral Importance*, pp. 215–219; Schrader-Frachette, "Parfit and Mistakes in Moral Mathematics"; Eggleston, "Does Participation Matter?"; and Petersson, "The Second Mistake in Moral Mathematics Is Not about the Worth of Mere Participation."

16. Brandt, "Some Merits of One Form of Rule Utilitarianism," p. 133, n. 15.

17. Lyons, *Forms and Limits of Utilitarianism*, pp. 128–132 and pp. 137–142.

18. See Mackie, "The Law of the Jungle"; and Mackie, "Cooperation, Competition and Moral Philosophy."

19. See my *Ideal Code, Real World*, pp. 164–169.

20. Murphy, *Moral Demands in Nonideal Theory*.

21. Feinberg, *Social Philosophy*, p. 111; and Wiggins, "Claims of Need," p. 10.

22. If getting wrinkled caused you to lose your spouse or your job, things would be different. I am assuming that you do not have the sort of spouse or job that you would lose because of wrinkles.

23. An influential advocate of the view that needs are what one must have satisfied if one is to have a decent life has been David Miller. See his *Principles of Social Justice*, especially p. 212 and p. 319, n. 23 and n. 25.

24. See my "Fairness, Needs, and Desert."

25. I discuss powerful arguments *against* weighted prioritarianism in *Ideal Code, Real World*, pp. 60–65, and in "When Is Impartiality Morally Appropriate?" at p. 39.

26. For relevant discussions, see the essays collected in Pojman and McLeod, *What Do We Deserve?* See also Kagan, *The Geometry of Desert*.

27. Rawls, *A Theory of Justice*, § 17 and § 48. For a helpful discussion, see Olsaretti, "Distributive Justice and Compensatory Desert," pp. 196–197.

15 Utilitarianism and the ethics of war

William H. Shaw

War has an obvious impact on human well-being, one that is almost always deleterious. Whatever good a given war might conceivably produce or whatever evils it may forestall, by definition it involves death and destruction, mayhem and misery. Even in periods of peace, the need to be prepared for war also affects human well-being, diverting human and material resources along channels that, viewed by themselves, are less productive of well-being than obvious alternatives to them. It is not surprising, then, that Jeremy Bentham and James Mill were concerned with the causes of war and how best to avoid it. However, neither they nor John Stuart Mill, who wrote about some of the violent conflicts of his day, addressed in sufficient detail two key ethical questions: (1) when, if ever, are we morally justified in waging war and (2) if recourse to arms is warranted, how are we permitted to fight the wars we wage? Nor have contemporary utilitarians examined these questions with the care they deserve. After reviewing briefly what Bentham and the Mills had to say about war, this essay addresses these two questions from a utilitarian perspective.[1]

Bentham and the Two Mills

Bentham disapproved of war and sought its elimination. "All war is in its essence ruinous," he wrote; "mischief upon the largest scale."[2] An anachronism in the modern world, war damages the interests of the masses, forcing them "to murder one another for the gratification of the avarice or pride of the few."[3] The few, in turn, rarely suffer the miseries brought about by the wars they direct. Bentham firmly rejected the idea that war benefits the national economy. It leads, rather, to higher taxation and increased executive power, and the colonies and trading privileges that are sometimes won by war do little or nothing to increase a nation's wealth. On the other hand, he believed that the frequency of war could be reduced by free trade, the development of international law, a foreign policy that was open to scrutiny and based upon non-interference with other nations, and by an effort to combat popular enthusiasm for war by debunking concepts like national honor and martial glory.

Bentham was not an out-and-out pacifist, however, because he believed that genuinely defensive wars could be justified. Noting in *Principles of International Law* that all states regard themselves as bound to protect their subjects against injuries from the subjects or governments of other states, he wrote that "the utility of the disposition to afford such protection is evident."[4] In that essay, he goes on, in his characteristic systematizing way, to catalogue various types of war, to review their typical causes or triggers, and to suggest some means for preventing them. For example, defensive

confederations can reduce or eliminate wars resulting from fear of conquest.

Suppose a state sees itself as aggrieved or finds its rights violated. Is it reasonable to go to war against its aggressor? That will depend in part on the "state of his [the aggressor's] mind with relation to the injury." If there is no bad faith (*mala fides*) on his part, then "it can never be for the advantage of the aggrieved state to have recourse to war, whether it be stronger or weaker than the aggressor."[5] Whatever the injury in question, the expense of war will always outweigh it. However unjust the aggression may appear, it is better, Bentham believes, to submit to it than to encounter the calamities of war. Even if the aggressor is acting in bad faith, whether recourse to war is worthwhile will depend on the circumstances. Unless the aggressive attack represents the first step toward national destruction, prudence may well dictate yielding.

In *Principles* Bentham outlined "A Plan for an Universal and Perpetual Peace" based on two propositions: reducing the armed forces that any European nation may possess and emancipating all "distant dependencies." To promote its own welfare, he contended, Britain should give up its colonies and found no new ones. It should avoid all military alliances and any trade treaties intended to exclude other nations, and it should reduce its naval forces to the minimum necessary to defend its commerce against pirates.

Early in his career, James Mill was not particularly critical of war. At one point, he favored renewed hostilities against France and volunteered to defend the capital if an invasion came.[6] Over time, however, he grew increasingly opposed to war, coming to share many of Bentham's views – for instance, that war damages the national economy and brings suffering and ruin to the masses:

> To what baneful quarter . . . are we to look for the cause of the stagnation and misery which appear so general in human affairs? War! is the answer. There is no other cause. This is the pestilential wind which blasts the prosperity of nations. This is the devouring fiend which eats up the precious treasure of national economy, the foundation of national improvement, and of national happiness.[7]

On the other hand, Mill joined Bentham in believing that wars of self-defense and possibly even preventive wars might be justified.

In his essay "Law of Nations," James Mill argued for (1) developing a clear code for identifying the rights of nations and regulating their conduct with one another and (2) establishing a tribunal for executing that law promptly and accurately. Nations should have recourse to war only when some right has been violated, when that violation is a serious one, and when remedying it requires the extreme measure of war. Further, the only just ends of war are compensation for the injury received and security against any fresh injury, and the law of nations should outlaw any violence not conducive to those ends. Consistent with this, Mill held that although combatants may attack other combatants, they are not justified in inflicting harm on them beyond what is necessary to

take them out of the fight. Thus, belligerents should take prisoners whenever possible, and they should treat them humanely. When it comes to civilians, "forbearance and preservation" rather than destruction "ought to be the rule . . . only to be infringed upon special and justifying circumstances."[8] The advantage that an invading army can gain from seizing and destroying the property of ordinary citizens "bears, unless in certain very extraordinary instances, no sort of proportion to the evil inflicted upon the individuals" and should therefore be forbidden by the law of nations.[9]

For his part, John Stuart Mill seems never to have discussed the causes of war or how to avoid it, and he said little about the ethics of war in general. However, in his discussions of current events, he occasionally made some pertinent remarks. In "A Few Words on Non-Intervention," for instance, Mill argued that nations should not take sides in civil wars or uprisings against an established government except to counterbalance the intervention of other outsiders.[10] When the people are struggling against their own government for free institutions, there is obviously no question of supporting the government against them. On the other hand, if an oppressed people are to win enduring liberty, Mill maintained, then they must do so themselves. The situation changes, however, if they are struggling against foreign rule or a domestic tyranny supported by outsiders. Then the reasons for non-intervention no longer apply.

Mill followed the American Civil War closely. He was scathingly critical of the South and hoped that the war would last long enough, as in fact it did, to become unequivocally anti-slavery in character. In the context of rebuking those who wished that the North had not resorted to arms or who were pushing for it to come prematurely to terms with the South, he remarked:

> [W]ar, in a good cause, is not the greatest evil which a nation can suffer. War is an ugly thing, but not the ugliest of things: the decayed and degraded state of moral and patriotic feeling which thinks nothing *worth* a war, is worse . . . A man who has nothing which he is willing to fight for, nothing which he cares more about than he does about his personal safety, is a miserable creature . . . As long as justice and injustice have not terminated *their* ever renewing fight for ascendancy in the affairs of mankind, human beings must be willing, when need is, to do battle for the one against the other.[11]

Mill discussed rebellion and insurrection in various other contexts as well.[12] When was such political violence justified? It was defensible, he thought, only if it had a just cause. For Mill, this involved suffering, oppression, or tyranny, the intensity of which was so great that overcoming it was worth "almost any amount of present evil and future danger."[13] Even with a just cause, however, for rebellion or insurrection to be warranted, it must also have a reasonable chance of success.

When, if ever, may we wage war?

It is extremely unlikely that any war, viewed as a whole, has been a welfare-optimal event or series of events. There will almost certainly have been some alternative state of affairs that the nations in question could have brought about instead that would have been better on utilitarian grounds. Indeed, there will often have been an outcome open to the belligerents that would have been better for each of them individually. That's why wars so often look like collective folly. However, the decision that an individual state faces is not how all states should act, but rather how *it* should act in the particular situation in which it finds itself.

When, then, should a state go to war? From a utilitarian perspective, it is not enough that the war's benefits outweigh its costs, taking into account the interests of all. There must also be no alternative course of action open to the state that would lead to better results (for example, not responding to the provocation, conceding land or influence, surrendering, or relying on nonviolent, civil resistance). Formalizing this idea and filling in a few details, we can say that utilitarianism entails the following principle:

> A state is morally justified in waging war *if and only if* no other course of action available to it would result in greater expected well-being; otherwise, waging war is wrong.

This principle is formulated in terms of *expected* outcomes. Some utilitarians, however, prefer to frame the theory in terms of *actual* outcomes. There are considerations for and against both approaches.[14] In practice, though, there is little difference between them because the only way we have of trying to maximize actual well-being is to act so as to maximize expected well-being.

If a state is justified – if it acts rightly – in waging war, then it is morally permitted to do so. Is it also required to wage war? The answer is that a state is morally required to wage war, as opposed to being merely permitted to do so, if and only if waging war would result in greater expected well-being than anything else it could do. This implies that, if the well-being expected from some other course of action (A) is equivalent to that of waging war (W), and no course of action B, C, D, . . . etc. has equivalent or greater expected utility, then although a state must choose either A or W, it is not required to select one rather than the other. It acts rightly if it opts for either. This may sound like a purely theoretical point. But given the difficulties of estimating expected well-being, the possibility of ties may not be so remote.

Two further points should be clarified. First, there are various kinds of war that states can fight. This entails that a state will often be presented not simply with a choice between waging war and not waging war but rather between the latter and various possible wars of differing scope, type, or intensity. Second, the decision whether to wage war is not a one-time, once-and-for-all decision. Even if war is initially warranted on

utilitarian grounds, circumstances can change so that a state ceases to be justified in continuing it.

The utilitarian approach to war is simple, but it is also attractive. Commonsense morality, I believe, would recoil at the suggestion that it can be right to wage war when there is an alternative to fighting that would have better results. And if well-being is understood in a broad sense that acknowledges the importance of liberty and self-determination for human flourishing (as Mill and many subsequent utilitarians have thought), then commonsense morality might well accept the explicitly utilitarian proposition that wars should be waged only when nothing else would bring about greater well-being. One does not have to be a utilitarian, however, to believe that this normative standard is basically correct. One can consistently endorse it while nevertheless believing that utilitarianism does not provide a fully satisfactory general account of right and wrong. A non-consequentialist, that is, could well believe that utilitarianism is deficient as an overall normative theory, perhaps because it sometimes demands too much of us or fails to place sufficient weight on our individual projects and attachments, and yet believe that wars ought to be waged only when doing so satisfies the utilitarian standard.

Widespread acceptance of the utilitarian criterion of justified war would be a profound and salutary change. Throughout history, wars have been defended in terms that either ignore consequences altogether or focus on consequences that have little or nothing to do with human well-being, that is, with how well or poorly real people will fare as a result of war. For example, wars have often been defended as necessary to uphold national honor or avenge some historical grievance. And even when consequentialist-sounding considerations are adduced in support of war, policy-makers typically discount or neglect altogether the consequences for other peoples, and they rarely think through the probable gains and losses of fighting in a detailed and objective way. They may claim, and frequently do, that there is no viable alternative to war, that the stakes are extremely high, and that the consequences of not fighting would be too terrible to contemplate, yet these claims are almost always overconfident and unsubstantiated.

Three common objections addressed

Many of the common objections to utilitarianism as a comprehensive moral theory are inapplicable to its criterion of justified war, which is a relatively specific normative principle of limited domain. However, there are three objections to utilitarianism that are widely seen as having special force when applied to its treatment of war. Let us examine these.

The first objection is that utilitarian reasoning is susceptible to abuse – in particular, that it can be, and has been, used to justify all sorts of immoral wars. To be sure, wily leaders have often rationalized the wars they wished to fight by spurious appeals to the greater good. However, they have not refrained from also exploiting various non-

utilitarian rationales – appealing, for example, to religious duty, to the glory or honor of the nation, or to some historical injustice it has allegedly suffered. The devious or deluded can twist almost any moral principle into supporting morally objectionable policies. Utilitarianism is no more prone to misuse of this sort than are other moral theories. Even if it were, this would not show that it fails to provide a satisfactory normative criterion for assessing wars, only that the criterion has to be handled with care. Furthermore, the proper response to superficial and fallacious consequentialist arguments for war is not to abandon consequentialist thinking to the ignorant and unscrupulous, but rather to examine with as much specificity and meticulousness as possible the likely results of waging war in the given circumstances, taking into account the interests of all.

But what are those interests? Utilitarians wish to promote the well-being of sentient creatures as much as possible. For them, well-being is all that ultimately matters. The second objection, however, is that well-being is an uncertain concept, and that it is hard to know what really does make people's lives go better or worse. Now it is true that philosophers have different conceptions or definitions of well-being and also that, whatever conception of well-being one favors, there is much we do not know about the social and psychological factors that conduce to human flourishing. However, even if we have a lot left to learn, we already know quite a bit about the things that promote well-being, on the one hand, or impair or thwart it, on the other. Moreover, when it comes to war, we are not dealing with subtle questions of value. Killing and maiming people, orphaning children, destroying farms and factories, dislocating civilians, ripping up the economic infrastructure of a country and damaging its cultural heritage – these are patently destructive of well-being. On the other hand, there is no question that physical and psychological security, personal liberty, political self-determination, and respect for territorial integrity promote well-being and that war may sometimes protect these values.

But even if we can make plausible assessments of comparative value, our knowledge of the future is tenuous and uncertain. This is the basis of the third objection, namely, that when it comes to specific questions of war and peace, the future is too unpredictable for us to make reliable judgments about the likely consequences of the alternatives open to us. It has to be granted that, when making decisions, people can and frequently do overlook some courses of action altogether, and with regard to the alternatives they do consider, they often miss some possible outcomes or miscalculate their probability. Further, human reasoning is often plagued by faulty background assumptions, wishful thinking, and various cognitive errors. And especially when it comes to war, emotion and social pressure can further cloud people's judgment. Our epistemic situation is far from hopeless, however. We often foresee reasonably well where our actions will lead, and our anticipations of the future frequently prove more or less correct. When it comes to war, in particular, it may be possible to learn from the past. At the very least, history can teach us to be more modest and circumspect and to acknowledge our own fallibility.

Hardly anyone has used the utilitarian standard to assess particular wars. If they were to do so, it seems clear that it would imply that many wars were morally unjustified and

should not have been fought and that neither side acted in a way that promoted the well-being of all. About some wars, it may not be obvious what careful utilitarian reflection would show. I shall consider below some strategies for making the utilitarian criterion easier to apply in practice, but when it comes to big, complicated real-world questions of war and peace, we have no reason for assuming that the correct normative principle, whatever it may be, will always give us simple and effortless answers. And, as suggested earlier, the fact that an ethical criterion is sometimes, or even often, difficult to apply does not show that it is incorrect.

Intermediate principles

In *Utilitarianism*, Mill speaks of the importance of following "intermediate generalizations" or "corollaries from the principle of utility" instead of "endeavour[ing] to test each individual action by the first principle." He continues:

> It is a strange notion that the acknowledgment of a first principle is inconsistent with the admission of secondary ones. To inform a traveller respecting the place of his ultimate destination, is not to forbid the use of landmarks and direction-posts on the way . . . Whatever we adopt as the fundamental principle of morality, we require subordinate principles to apply it by.[15]

Applied to the context of war, Mill's remarks imply that even if utilitarianism provides the correct criterion for distinguishing justified from unjustified wars, the theory itself, recognizing the difficulty of applying that criterion accurately, may recommend that policy-makers follow certain secondary or lower-level principles. Doing so may reduce the epistemic and other difficulties that can impede employing the utilitarian criterion successfully.

What principles might successfully play this role? Two possibilities suggest themselves: pacifism and the principles of just war theory. Let's consider these in turn.

(a) the pacifist principle. By this I mean the tenet that it is always wrong to wage war. Any humane person will want people to believe, and to be disposed to act on the belief, that war is a dreadful thing and should be strenuously avoided. The pacifist principle, however, goes further than this, affirming categorically that a state should never wage war. Why might utilitarians want people to follow this principle? They might desire this if they believed that waging war is never – or only extremely rarely – the optimal course of action. If so, this could then lead them to a kind of sophisticated pacifism, which holds that although in theory war might be morally justified, in fact this happens so rarely that we will get the best results if we internalize in ourselves, teach our children, and proclaim to others the principle that war is categorically wrong and that no nation should ever again fight one. Even if adhering to this pacifist principle leads us on rare occasions to err morally by failing to fight wars that the utilitarian standard requires us to

fight, we still do better in the long run, the argument continues, by cleaving to pacifism and refusing ever to entertain war as an option. This is because the danger of misapplying the utilitarian criterion and fighting wars that are morally wrong dwarfs the possibility that by sticking to the pacifist principle we will neglect to fight wars that we ought to fight.

Although clever and provocative, this argument is ultimately unconvincing, I think. It might seem obvious that a world in which everyone believed that wars should never be fought would be a better world because no wars would be fought. But whether that is true depends on what the imagined world without war would be like. If such a world had more injustice, oppression, or aggression – in short, more misery and less well-being – than a world with the occasional war, then it might not be a better world. Furthermore, to follow the sophisticated pacifist and repudiate all war may or may not be the best way to bring it about that human beings fight no wars. It depends on whether most other people also categorically repudiate war. If they have trouble accepting pacifism or have trouble sticking to it when push comes to shove, advocating it might not be the best strategy for lessening the frequency of war, let alone ridding the world of it altogether.

Suppose, what seems entirely plausible, that in the foreseeable future the vast majority of people are not likely to be brought to believe that we should never, ever wage war regardless of the circumstances. Imagine, then, that a nation finds itself in a situation similar to those that in the past have provoked nations to fight. There will certainly be people arguing for going to war on various grounds, moral and otherwise, sound and spurious. Suppose further that it would indeed be wrong to fight the war in question. How is that war to be effectively argued against without looking at the facts, that is, without bringing a utilitarian perspective to bear? Simply to reaffirm one's categorical opposition to war, as our sophisticated pacifist does, and to say "we don't have to look at the details to know that this war would be wrong" will be unconvincing when the war party is making arguments, giving reasons, and pointing to facts. By requiring us to focus on the human costs and benefits of any possible war, to examine likely outcomes, and to weigh alternative courses of action, utilitarianism invites decision makers to be specific and empirical. If a prospective war is indeed wrong (as we are now supposing), then arguments that focus on consequences and alternatives are likely to be more effective in opposing it than is refusing to debate the particulars of the situation with the proponents of war. Most countries face the decision whether to wage war only at infrequent intervals when the circumstances seem exceptional, when what is at stake appears very important, and when emotions run high. It is precisely here that the absolutist rejection of war that the sophisticated utilitarian-minded pacifist wants us to teach is likely to be ineffective and to be set aside as only a generalization open to exceptions (which, of course, is exactly what it is for our sophisticated pacifist).

(b) the principles of just war theory. Another, more plausible possibility is that utilitarians should recommend following the principles developed over the centuries by various thinkers in the just war tradition as specifying when it is morally legitimate to

wage war. Specifically, just war theory holds that a war may permissibly be fought if and only if (1) it has been authorized by a legitimate authority, (2) it has a just cause, (3) it is undertaken with the right intention and (4) as a last resort, (5) the harm the war will do is not out of proportion to the good it will achieve, and (6) there is a reasonable prospect of success.[16]

The proposal, again, is that utilitarians should adopt these principles as intermediate principles in Mill's sense. Although ultimately subordinate to the utilitarian criterion, they would provide the workaday framework in which the morality of resorting to war is to be examined. Somewhat easier to work with than the utilitarian criterion itself, these principles draw our attention to considerations that are clearly relevant from a utilitarian perspective. This is obvious for principles such as last resort, proportionality, or reasonable prospect of success, but one can also discern a consequentialist rationale for others, such as right intention and just cause, that appear on their face thoroughly non-consequentialist. For instance, although intention in and of itself is irrelevant to utilitarian analysis, it is doubtful that a government that is not sincerely trying to act on the basis of moral principle will end up waging a war that is morally permissible according to the utilitarian criterion. Thus, examining a prospective war with the aid of the just war guidelines reduces the likelihood that in deciding either to wage or to refrain from war we will contravene the utilitarian criterion. Furthermore, these six principles are widely known, and many people find them intuitively plausible. Cadets and midshipmen encounter them at the US service academies.

If utilitarians see practical merit in employing the just war principles, how exactly should they understand these rules? One possibility is to consider them as merely pragmatic guidelines or "rules of thumb"[17] that assist in the application of the utilitarian criterion but that lack any normative weight in their own right. Rules of thumb assist our decision-making, but are rightly put aside, without compunction or regret, when one thinks one is in circumstances to which they do not apply or in which reliance on them would lead one astray.

A second possibility, however, is to treat these secondary rules as genuinely moral in character, as having normative force. When a person has internalized a rule as part of his or her moral code, the person does not look at the rule instrumentally – it is not something the person can pick up or put down as the occasion demands. Rather, the person will tend to feel guilty when his or her conduct fails to live up to the rule and to disapprove of those who act contrary to the rule. This way of thinking about rules does not necessarily entail a rule-utilitarian criterion of right.[18] Sophisticated act utilitarians can agree that in some situations having people strongly inclined to act in certain rule-designated ways, to feel guilty about failing to do so, and to use that standard to assess the conduct of others can have enormous social utility.[19]

Is this one of those cases? To answer this question, one would have to determine whether policy-makers for whom the just war principles have normative force would be

more likely to act in ways that conform to the utilitarian criterion than would those who view them as only rules of thumb. This is an issue that needs further discussion.[20]

Three alleged counterexamples

When non-consequentialists criticize utilitarianism and other consequentialist normative theories, a standard technique is that of counterexample. The critic tries to imagine some scenario where utilitarianism implies a course of conduct that conflicts with our moral intuitions and then, on this basis, concludes that the theory should be repudiated. When it comes to war, however, it is difficult to find any real-world examples to play this role – that is, to find some actual war about which utilitarianism implies, contrary to commonsense morality, that a state either should or should not have fought it. Perhaps, though, there are plausible hypothetical examples where utilitarianism has implications that would strike one as counterintuitive or seriously problematic. Let's consider three possibilities.

One alleged counterexample focuses on the possibility that utilitarianism might authorize a war that lacks a just cause. More specifically, it proposes that a large nation, whose economy – and thus whose citizens' well-being – depends on oil, would be justified on utilitarian grounds in seizing the oil reserves of a small nation because doing so would maximize benefit.[21] This is extremely fanciful. We must imagine that the small nation flat-out refuses to sell its oil at any price, that the larger nation has no other way of obtaining oil, and that no one foresaw this situation developing in time to avoid it.

Even granting this scenario, utilitarians have good reason to insist that wars should be fought only for a just cause – in this case, for insisting that states should respect the territorial integrity and political sovereignty of other states (assuming they are not tyrannical or genocidal). Already entrenched in international law, this is a political and moral right that it makes sense for utilitarians to uphold because of its importance for promoting human well-being in the long run. But having endorsed and advocated that right, utilitarians cannot then turn around and authorize its violation in an effort to capture extra utility. Practically speaking, it is impossible to institutionalize a right (and reap the benefits that this brings) and at the same time license violations of the right in particular situations. Permitting states to ignore the basic rights of other states whenever they believe that doing so would maximize utility would be a disastrous policy.

A second alleged counterexample imagines that a large aggressive nation wrongfully invades a small nation. If resistance would be futile, then utilitarianism implies, it seems, that the small nation would be wrong to resist.[22] This strikes some people as counterintuitive. Doesn't a nation have a right to resist even if resistance is ultimately hopeless? To make the case more extreme, suppose the larger nation intends to exterminate the people of the smaller nation. Don't they have a right to go down fighting? The short answer is that, yes, nations have a right to use necessary and proportionate

force to defend themselves against armed attack, a right that it makes perfect sense for utilitarians to uphold, given the importance of respecting the political sovereignty and territorial integrity of nations.[23]

One can, of course, sometimes act wrongly by exercising a right as when, for example, I utilize my right of free speech to say hurtful things to you. But part of the point of treating something as a right in the first place is to enable people to act in certain ways without being obliged to calculate the utility of their doing so. Thus, utilitarians support the right of free speech because of the myriad ways in which upholding that right contributes to the human good in the long run, even if people's exercising that right fails to maximize utility in some cases. Furthermore, to return to our example, it is by no means obvious that it will do no good for the small nation to resist. Even apparently hopeless resistance may accomplish something. It might force the aggressor nation to modify its immediate plans or dissuade it from undertaking similar campaigns in the future. Given this possibility, it is unlikely that one could plausibly criticize the victim nation for trying to stave off the worst. Moreover, we often admire people who fight for a noble cause even when they are doomed to lose.

As a general matter, however, we want to discourage nations from continuing to fight after it becomes clear they will lose. And while we may admire a subjugated people for taking up arms against its oppressor despite the certainty of defeat, we are unlikely to encourage the next subject nation to do the very same thing. Fighting to the last man may sound romantic, but utilitarians and all humane people disapprove of pointless loss. But, again, suppose you will die anyway: Shouldn't you try to kill as many aggressors as possible before you are overwhelmed? Again, neither utilitarianism nor commonsense morality is likely to criticize you for exercising a right (the right of an individual or a state to defend itself) that it is essential to uphold. On the other hand, however, there can be dignity in submitting to an unjust fate peacefully and with your head held high – going to your execution calmly, for example, rather than struggling with your guards in a desperate attempt to inflict some final injury on them.

A third alleged counterexample takes off from the idea that it is wrong for a nation to fight a war if there is a better course of action open to it. Suppose, then, that from a utilitarian perspective it would produce more good for a nation to fight a certain war than to refrain from fighting and to carry on as usual. Suppose further, though, that there is a third thing it could do instead, say, launch an all-out drive to eradicate malaria in the third world, which would produce even more well-being than waging war. Utilitarianism implies that the nation should combat malaria rather than wage war.[24] This strikes some people as counterintuitive: If waging war is the right course of action when compared simply to not fighting, then how can that judgment change because of something that bears no intrinsic connection to the war or the issues it is about?

To begin with, the imagined scenario requires many improbable assumptions: for example, that the nation in question cannot both fight the war and mobilize against

malaria; that it never previously realized the good it could accomplish by trying to eradicate malaria; and that the welfare cost of deferring the malaria campaign until after the war would be too high. But let us grant the imagined scenario. Still, the war in question might involve the nation's right to defend itself against attack, or it might be required to wage that war by treaty obligations or considerations of justice. If so, then utilitarianism may concur with commonsense morality that waging war takes priority over aiding other nations to overcome malaria – that is, that it is more important for states to do what justice requires or for them to be allowed to exercise their basic rights than it is for them to aid less advantaged neighbors or act in other supererogatory ways. Suppose, though, that this is not the case. Then, utilitarianism does indeed imply that the state in question should campaign against malaria rather than wage war. Although a surprising conclusion, perhaps, it is not obviously counterintuitive. Ordinary Americans sometimes say, "We shouldn't be fighting over there when there is so much to be done at home." Although our example concerns helping other nations rather than promoting domestic well-being, the quoted remark suggests that everyday morality is untroubled by the idea that a war might be wrong because it prevents us from doing something that would produce more good.

Utilitarianism and the rules of war

Let us turn now to the question of what an individual is to do, once a conflict breaks out. If the war is morally justified on utilitarian grounds, then one should support it. If not, then – some special circumstances aside – one should oppose it. As the great utilitarian thinker Henry Sidgwick wrote, the duty of an individual whose nation is contemplating an unjustified war is clear; it is "to use any moral and intellectual influence he may possess – facing unpopularity – to prevent the immoral act." But "how far he should go in such opposition" is "difficult to say" and "depends so much on circumstance that an abstract discussion of it is hardly profitable."[25] It is even more difficult to say, in general, what an ordinary soldier should do if he comes to believe that he is fighting in an immoral war. Naturally, he will be reluctant to participate, but he may have conflicting moral, legal, and professional responsibilities. I shall not pursue this issue here.

Suppose, however, that the war is justified and that one is either a combatant or an official involved in directing the war effort. How, according to utilitarianism, is one permitted to fight that war? What constraints, if any, are there on one's conduct? It might seem that if the war is morally right, that is, if waging it maximizes expected well-being, then those involved in the fighting should be prepared to do absolutely anything as long as it helps to achieve victory.[26] This is erroneous for two reasons.

First, utilitarians seek to maximize net expected benefit; they care not only about how much good a course of action can be anticipated to produce, but also about the costs of bringing that outcome about. The fact that victory, viewed by itself, would be a better

result than, say, surrender or negotiation does not entail that it is right to fight in any way that conduces to that victory. If the harm of fighting in some way is too great, then the utilitarian calculus will entail that the war cannot be fought that way, even if this makes victory impossible.

The second reason utilitarians do not believe that combatants and those who direct them are permitted to do absolutely anything to obtain victory, even if utilitarianism says that they are in the right, is that their conduct is restrained by what I shall call the "recognized rules of war." These rules are of two sorts. One major component is the international law of armed conflict, which is based on customary practice and a wide variety of treaties and agreements, such as the Geneva Convention, which the nations of the world have ratified over the past 150 years. These rules are extensive and detailed. For example, there are rules outlawing certain weapons (like poison gas), rules concerning the treatment of the enemy sick and injured, and rules governing the handling of prisoners of war and the population of an occupied country. There are also rules that forbid, for example, taking hostages, declaring that no quarter will be given, abusing a flag of truce or Red Cross emblem, and soldiers' dressing in civilian clothing.

In addition, almost all writers on the ethics of war believe that combat is (or ought to be) regulated by three normative principles. These are the principle of necessity (which forbids violence that serves no legitimate military purpose), the closely related principle of proportionality (which holds that the use of force or violence must be proportionate to the value of the military objective being sought), and the principle of discrimination and civilian immunity (which requires combatants to distinguish between legitimate and illegitimate targets, to refrain from targeting civilians, and to minimize collateral harm to them). These three principles are reflected in various provisions of the law of armed conflict, but their normative validity is generally thought to be independent of it. Because both the laws of war and these three principles are widely recognized and endorsed (if not always respected in practice), I refer to them together as the recognized rules of war.

Utilitarianism clearly, staunchly, and unequivocally supports the recognized rules of war because adherence to them, even those rules that seem purely conventional or even arbitrary, reduces the carnage of war. From a utilitarian perspective, adherence to the rules is so important that even those who believe that act utilitarianism provides the ultimate criterion of rightness can agree that they should be treated not as mere guidelines or rules of thumb but, rather, as firm moral rules that should make up part of the personal and professional code of combatants. Although simple and, in a way, obvious, a utilitarian approach to the rules of war captures well why they matter so much. It is also easier to square, than are some other ways of thinking about those rules, with some of their distinctive and otherwise troubling features – in particular, the fact that the rules have to be such that states can reasonably be expected to endorse them; the fact that they apply to both sides equally whether they are right to fight or not; the fact that aspects of the rules seem somewhat arbitrary from the moral point of view (for example, the outlawing of poison gas but not flamethrowers or the fact that some combatants may

be morally innocent); and the fact that the rules, while trying to minimize the violation of civilian rights, do tolerate their violation.

Utilitarians and other humanitarian-minded people may seek to refine, clarify, or modify the recognized rules or even to introduce entirely new ones in an effort to make the rules as welfare-promoting as possible. In Henry Sidgwick's words, utilitarians seek that "system of rules of international conduct for which it is desirable to obtain – and not unreasonable to hope – general acceptance."[27] In other words, they seek those rules of war that, given the world as it is and nations and people as they are, will bring about the most good – or, rather, salvage as much well-being as possible in the dreadful circumstances of war – taking into account, among other things, the likelihood of belligerent nations and the people who fight for them being brought to accept and comply with those rules. For the foreseeable future, however, utilitarians will be even more concerned with trying to see that the recognized rules are subscribed to as widely as possible, that potential combatants and their leaders are taught the rules and internalize a commitment to them, that more military organizations make adherence to them part of their organizational culture, and that mechanisms are found for institutionalizing and enforcing them.

Civilian immunity

Let us consider more closely the very important civilian-immunity rule. From a utilitarian perspective, it is desirable to affirm and, as far as possible, entrench the right of civilians not to be attacked and to oblige warring states to take whatever steps they reasonably can to avoid injuring them or their property. Utilitarians want states and quasi-state actors to recognize, uphold, and institutionalize this right; to have military establishments pledge their allegiance to it and to train their troops to follow it; and to have individual soldiers internalize a commitment to the rule as part of their professional code.

Some philosophers, however, have doubted that utilitarians really can endorse civilian immunity without reservation, in particular, that they can ratify a categorical ban on directly targeting civilians. Douglas Lackey, for instance, maintains that utilitarianism entails that belligerents should simply choose the means of attacking the enemy that will cause the fewest deaths and injuries, counting combatant and civilian casualties as equivalent.[28] Because the utilitarian goal is harm reduction and because the lives of civilians are neither more nor less valuable than those of combatants, if attacking civilians would result in fewer deaths overall than would attacking combatants only, Lackey reasons, then that is the correct utilitarian course of action. No doubt, one can imagine hypothetical scenarios in which this might seem plausible. But utilitarians will consider the matter in broader terms than Lackey thinks. To abandon the civilian-immunity rule and thus permit belligerents to target civilians whenever they think that doing so would be for the greater good would clearly have dreadful results. The direct and indirect civilian

casualties of war are often shockingly high,[29] and one of the few things the world can do to reduce those casualties is to continue insisting that belligerents always discriminate between combatants and non-combatants.

If utilitarians looked at civilian immunity in the narrow, case-by-case way that Lackey suggests, then for them it would be, at best, only a rule of thumb. Although it usually promotes utility to respect civilians, their immunity should be put aside if the circumstances dictate doing so. I have been arguing that this is the wrong way for utilitarians to view the rules of war. But even if those rules are treated not merely as useful guidelines, but rather as rules that people feel morally obliged to follow, some utilitarians, such as Richard Brandt, appear to believe that they should be formulated so as to permit direct attacks on civilians in certain circumstances.[30] Many non-utilitarians also believe this. For instance, Michael Walzer allows that sometimes the rule of civilian immunity must yield to direct utilitarian calculation. More specifically, he believes that in certain special circumstances, which he calls a "supreme emergency," a state may be morally justified in directly attacking civilians if this is the only way to preserve a political community that would otherwise be destroyed.[31]

Walzer's stance faces problems as does the less developed position of Brandt.[32] First is the difficulty of specifying the exact circumstances that constitute a supreme emergency and thus permit a belligerent to override civilian immunity. Second is the fact that states could easily imagine that they are in a supreme emergency when they are not. Third is the precedent effect. Even if a nation acts rightly in violating civilian immunity, doing so sets a bad precedent, making it likely that in the future both it and others will break the rule when they should not.[33] For example, even if, as Walzer believes, Britain faced a supreme emergency in the early years of World War II, it continued bombing German cities throughout the war, long after any such emergency had passed. Further, the policy paved the way for American bombardment of Japanese cities, culminating in Hiroshima and Nagasaki.

For these and related reasons, Stephen Nathanson has refused to follow Brandt or Walzer, arguing instead from a utilitarian perspective that the civilian-immunity rule should not be revised, nor should exceptions to it be permitted. Rather, utilitarians should endorse it categorically.[34] Still, some utilitarians are bound to worry that, whatever the rule is, there will inevitably be exceptions to it – cases where it would be better to violate the rule than to adhere to it. Let us look at this more closely.

Suppose, then, that some soldiers or those in charge of them believe that in their precise circumstances adhering to the civilian-immunity rule would jeopardize their ability to prevail. It cannot be denied that they might be correct, that it could be the case that violating the rule in that situation would be the utility-maximizing thing to do. Nevertheless, this is only a remote possibility. Even if we assume that directly killing civilians makes it possible to win a skirmish, battle, or campaign, it may contribute little

to ultimate victory. And even if it does, victory for their side, especially when pursued this way, may not be the optimal outcome. It is extremely likely that there will be alternative courses of action that are welfare-superior to a belligerent pursuing victory in defiance of the recognized rules of war.

Even if the soldiers in our example were in an exceptional situation where, all things considered, civilian immunity should be disregarded, they will almost certainly lack reasonable grounds for believing that they are so situated. The temptation to violate civilian immunity generally occurs in circumstances that are far from conducive to making reliable moral judgments – soldiers in the heat of battle or military strategists emotionally absorbed by the effort to prevail over another state. Indeed, it is hard to imagine worse epistemic conditions for making a balanced, objective, long-term assessment of the costs and benefits of breaking a rule, general adherence to which is so important.

Finally, the well-trained and conscientious soldier or war leader will not find it easy to violate the civilian-immunity rule even if he believes doing so would be for the greater good. That rule is part of his personal and professional moral code, and violating it will go against his conscience. He may even have taught the rule to others or criticized them for violating it. Moreover, as a utilitarian he approves of this situation; that is, he approves of his having a character structure that makes it virtually impossible for him intentionally to kill civilians. And he also knows that utilitarians will not blame or criticize him for adhering to a rule that they want to see instilled in combatants as deeply as possible.

No doubt, this is an issue that utilitarians need to explore further and more deeply. However, there is a strong case, I believe, for contending that belligerents should stick to the civilian-immunity rule unconditionally. Indeed, if there is ever a utilitarian rationale for treating a rule as, in practice, categorical and without exceptions, the civilian-immunity rule would seem to be such a rule. We get better results in the long run if military strategists and soldiers in battle never even entertain the idea of killing civilians. This could, hypothetically, lead them to forgo certain opportunities to maximize well-being by killing non-combatants. Even so, we achieve more good by insisting that they always follow this rule than we would from adopting any other stance.

Notes

1. The second, third, and fourth sections of this essay draw on Shaw, “Utilitarianism and Recourse to War.”

2. Bentham, *Principles of International Law*, p. 544 and p. 552.

3. As quoted by Conway, "Bentham on Peace and War," p. 87. This paragraph is indebted to Conway.

4. Bentham, *Principles of International Law*, p. 544.

5. Bentham, *Principles of International Law*, p. 545.

6. Yasukawa, "James Mill on Peace and War," pp. 179–184. This paragraph is indebted to Yasukawa.

7. J. Mill, *Commerce Defended*, p. 119.

8. J. Mill, "Law of Nations," p. 22.

9. J. Mill, "Law of Nations," p. 22.

10. J. S. Mill, "A Few Words on Non-Intervention," *Collected Works*, vol. XXI, pp. 109–124.

11. J. S. Mill, "The Contest in America," *Collected Works*, vol. XXI, pp. 141–142.

12. G. Williams, "J. S. Mill and Political Violence." This paragraph relies on Williams.

13. As quoted in G. Williams, "J. S. Mill and Political Violence," p. 110.

14. For further discussion, see Shaw, *Contemporary Ethics*, pp. 27–31.

15. J. S. Mill, *Utilitarianism*, *Collected Works*, vol. x, pp. 224–225 (chapter 2, para. 24).

16. For a brief account, see Shaw, "Just War Theory."

17. Smart, "Outline of a System of Utilitarian Ethics," p. 42.

18. Hare, *Moral Thinking*.

19. On the role of rules in utilitarian thinking, see Shaw, *Contemporary Ethics*, chapter 5.

20. For one perspective, see Shaw, "Utilitarianism and Recourse to War," pp. 397–401.

21. I thank Chris Eberle for pointing out this objection.

22. I thank Uwe Steinhoff for raising this point.

23. Shaw, "Consequentialism, War, and National Defense."

24. Jeff McMahan raised this objection in a discussion of the utilitarian approach to war at the Stockdale Center for Ethical Leadership's April 2010 McCain Conference titled "The Ethics of War Since 9/11," held at the US Naval Academy.

25. Sidgwick, "The Morality of Strife," p. 52.

26. Walzer, *Just and Unjust Wars*, pp. 133–134. Yasukawa believes that this was the view of James Mill ("James Mill on Peace and War," p. 191).

27. Sidgwick, *The Elements of Politics*, p. 238.

28. Lackey, *The Ethics of War and Peace*, pp. 64–65.

29. Nathanson, *Terrorism and the Ethics of War*, p. 202.

30. Brandt, "Utilitarianism and the Rules of War," pp. 156–160.

31. Walzer, *Just and Unjust Wars*, pp. 251–268.

32. For good critiques of Walzer and Brandt, see Nathanson, *Terrorism and the Ethics of War*, pp. 146–159 and pp. 195–205.

33. Cf. G. E. Moore, *Principia Ethica*, pp. 163–164; Austin, *The Province of Jurisprudence Determined*, p. 44.

34. Nathanson, *Terrorism and the Ethics of War*, chapters 14 and 15.

16 Utilitarianism and our obligations to future people

Tim Mulgan

Unless something goes drastically wrong in the next few centuries, most of the people who will ever live are yet to be born. Our actions have potentially enormous impact on those who will live in the future. Our decisions affect who those future people will be, and even whether there will be any future people at all. The threat of environmental crisis gives us some inkling of the magnitude of our potential impact on future generations. Only in the last few decades have moral philosophers really begun to grapple with the complexities of intergenerational ethics. Underlying their often technical debates are some of the deepest moral questions. What makes life worth living? What do we owe to our descendants? How do we balance their needs against our own?

For the utilitarian, our obligations to future people are perhaps the most important part of morality. If our goal is to maximize the happiness of sentient beings, then the happiness of future people is of paramount ethical concern. Two distinctive features of utilitarianism are especially salient here. First, utilitarians are committed to *impartiality*. In the famous phrase attributed to Jeremy Bentham: "everybody to count for one, and nobody for more than one."[1] On its most natural reading, this commitment includes *temporal* impartiality. Human well-being is equally valuable, no matter whose it is – or *when* they live. Second, again following Bentham, utilitarians are suspicious of *egoism*. We must guard against our natural tendency to give undue weight to our own interests, values, traditions, or perspectives; or to *believe* what suits our interests, aligns our duties with our inclinations, confirms our prejudices, or otherwise enables us to think well of ourselves. As a result, utilitarians are especially suspicious of moral principles that privilege our *present* interests.

This chapter explores influential utilitarian accounts of our obligations to future people, and also offers suggestions for the future. Unlike many areas of ethics, the study of our obligations to future people is in its infancy, and there is much room for progress. One significant – and sobering – development is the realization that we can no longer assume that each generation will always be better off than the last. The possibility that our current way of life may produce a broken world – where future people do not enjoy the background of abundance that we take for granted – gives intergenerational ethics both a new context and a new urgency. We are only beginning to work through the changes that utilitarianism must undergo to be relevant in such a world.[2]

The first section below presents utilitarian arguments against non-utilitarian accounts. The next section addresses a range of puzzles of aggregation – exploring the difficulties surrounding the evaluation of different possible futures. The next two sections explore ways that the need to take account of future people may impact on other debates within utilitarian theory, such as debates about best account of right action and the best account

of well-being. The final section asks how the prospect of a broken future might impact on utilitarianism.

How non-utilitarians fail future people

For the utilitarian, obligations to future people are theoretically unproblematic. They have the same basis as our obligations to present people – the fact that our actions impact on the well-being of sentient beings. The precise content and scope of our obligations is controversial, as we will soon see. But this creates puzzles *within* utilitarianism, rather than threats to its very coherence. By contrast, utilitarians argue that many non-utilitarians cannot even make sense of intergenerational obligations, for three principal reasons: the non-identity problem, the lack of intergenerational reciprocity, and misplaced optimism.

The non-identity problem

The first problem for non-utilitarian theories was made famous by Derek Parfit, whose *Reasons and Persons* set the scene for contemporary utilitarian discussion of future people. Parfit first distinguishes two kinds of moral choice: *same people* (where our actions affect what will happen to future people, but not who will exist) and *different people* (where our actions do affect who will exist).[3]

Classical utilitarianism treats same people choices and different people choices identically. What matters is how happy people are, not who they are. In Parfit's terminology, utilitarians endorse a *no difference view*.[4] If the only difference between two choice situations (A and B) is that A is a same people choice and B is a different people choice, then there is no moral difference whatsoever between A and B. Under the no difference view, different people choices do not – per se – present any new ethical issues.

Parfit argues that its ability to treat same and different people choices identically gives utilitarianism a prima facie advantage. Many non-utilitarian theories are designed for same people choices, and thus cannot cope with different people choices. This matters because such choices are more frequent than we think. Consider two simple tales introduced by Parfit.[5]

> *Mary's Choice*. Mary is deciding whether to have a child in summer or winter. Mary suffers from a rare medical condition. Any child she has in winter will suffer serious ailments. A summer child, by contrast, will be perfectly healthy. On a whim, Mary opts for a winter birth.
>
> (Despite his ailments, her child has a life worth living.)

Mary's behavior seems morally wrong. But why? Intuitively, Mary acts wrongly because she harms her child. But the winter child has a life worth living – and would not otherwise have existed at all. (A child born in summer would be a different person, because he or she would have a different genetic makeup.) How can someone be harmed if they would not otherwise have existed at all? (It would be even more odd to say that Mary harms the child she would have had in summer. How can you harm someone who *never* exists?)

> *Risky Policy*. We must choose between two energy policies. The first is completely safe. The second is cheaper, but riskier. Perhaps it involves burying nuclear waste where there is no earthquake risk for several centuries, but a significant risk in the distant future. Suppose we choose the risky policy. Many centuries later, an earthquake releases radiation, killing thousands of people.

Again, our choice seems clearly wrong. But why? Intuitively, we do wrong because we harm those who die. But suppose the two energy policies lead to radically different futures – with different patterns of resource development, migration, and social interaction. Now take any particular individual killed by the catastrophe. Suppose the precise chain of events leading to her existence would not have occurred if we had chosen differently – her parents would not have met, and might not even have existed themselves. But now it appears that we have harmed no one. For how can someone be harmed by an action without which she would not exist? And, if we harm no one, how can our choice be wrong?

Parfit also offers a less dramatic example, where we choose between depleting and conserving natural resources.[6] Suppose we opt for depletion – delivering a better life for ourselves, but leaving few resources for future generations. Now consider the position of those who live in this depleted future. Ex hypothesi, they are worse off than those who would have lived if we had chosen to conserve. However, these future people still have lives that are well worth living – and they would not have existed at all if we had chosen conservation. In this case, where no specific future catastrophe results from our decision, it is especially hard to construct a harm-based complaint on behalf of these future people. Yet still our choice seems wrong.

These examples illustrate the *non-identity problem*. This problem arises when, because different actions bring different individuals into existence, we have an action that seems intuitively wrong, but we cannot point to any particular individual who has been wronged or even harmed.

The non-identity problem brings together morality and metaphysics, as any claim that different people exist in different possible futures rests upon some theory of what makes each of us the person she is. While there are metaphysical controversies at the boundaries, almost everyone agrees that, for instance, a possible world where your actual

parents never meet is a place where *you* never exist. And this claim is sufficient for Parfit's examples. We cannot sidestep the non-identity problem by denying that different people choices are widespread.

Non-identity is a significant problem for any *person-affecting principle* that says an action can only be wrong if some particular person is worse off than they would otherwise have been. In a different people choice, whatever we do, no particular individual is worse off than she would otherwise have been, since she would otherwise not have existed. Utilitarians argue that no person-affecting principle can ever condemn any response to a different people choice – and therefore that no such principle is acceptable.

Non-identity puzzles abound in practical ethics, especially medical ethics. Any individual reproductive choice or medical procedure affects the identity of the resulting child. For instance, many people object that new reproductive technologies *harm* the resulting children. But, if they affect a child's genetic makeup, and if genetic identity is a component of individual identity, then such technologies involve different people choices – and the resulting children would not otherwise have existed. If their lives are worth living, how can they be harmed? Non-identity issues also arise in discussions of reparations for historical injustice. Should present people be compensated for some past wrong if they would not have existed in an alternative future where the injustice did not occur? Can the descendants of those who suffered from slavery or colonization consistently complain about injustices without which they themselves would never have existed?[7]

The non-identity problem is especially difficult for the social-contract tradition, where justice is modeled as a bargain or agreement among rational individuals. How can we begin to imagine contracts, bargains, or cooperative schemes involving future people whose existence and identity depend upon what we decide? Contractualists as diverse as Immanuel Kant, John Rawls, David Gauthier, and T. M. Scanlon all face serious difficulties here.[8]

Some non-utilitarians respond to the non-identity problem by biting the bullet. David Heyd, for instance, denies that we have any obligations *to* future people. All our obligations relating to the future are really owed to our contemporaries.[9] Most non-utilitarians, however, attempt to dissolve Parfit's non-identity problem. Whether their attempts succeed is a subject of ongoing controversy.[10] Much debate focuses on whether a non-utilitarian theory must be person-affecting.[11] Consider Scanlon's influential form of contractualism, where each agent asks what moral principles she and others could reasonably reject in some actual situation. This certainly appears to be a person-affecting theory, and opponents object that contractualism faces the non-identity problem. Contractualists seek to rebut this claim. For instance, Rahul Kumar reinterprets contractualism to include our obligations to individuals who occupy particular roles – or

fall under a given generic description – whoever those individuals might be.[12] In Kumar's own example, a woman considers her moral relationship to "my future child" – even though her present decision is identity-affecting for that very child.[13] Alternatively, Parfit incorporates impartial and impersonal reasons directly into contractualism – so that the interests of distant future people can provide present people with direct reasons to reject moral principles.[14]

Such reinterpretations blur the boundaries between utilitarian and person-affecting theories. One common worry is whether, to avoid the non-identity problem, non-utilitarians must abandon their distinctive commitments. Indeed, Parfit himself rejects Kumar's "generic persons" approach on the grounds that it abandons one of Scanlon's most appealing contractualist ideas – the thought that morality is something we owe to *individuals*.[15] But Parfit's own solution will strike many contractualists as a step too far – especially as he goes on to argue that his contractualism coincides with the best form of consequentialism.

The non-identity problem grounds a prima facie argument for utilitarianism. Because it adopts a no difference view, utilitarianism can easily say what is wrong with Parfit's risky policy, or with any action that creates people who are less happy than other people who might have existed instead.

The lack of intergenerational reciprocity

Another problem for non-utilitarian theories is the lack of reciprocal interaction between present people and distant future people. We can do a great deal to (or for) posterity, but, as the saying goes, what has posterity ever done for us? This is a problem because many non-utilitarians base morality on reciprocal interaction.

Imagine a time bomb that devastates people in the distant future but has no direct impact until then. (Real-life analogues might involve the storage of nuclear waste or the destruction of the global climate.) Suppose the people who suffer live so far in the future that no one alive today cares for them. Intuitively, it seems very wrong to gratuitously plant a time bomb. But how can any theory based on reciprocal interaction deliver this result?

Here we reach intuitive bedrock. Utilitarians advocate strict temporal neutrality. Planting a time bomb is just as wrong as planting a bomb that explodes today. Some non-utilitarians bite the bullet. Given that morality and justice require existence and interaction, then we have no moral obligations, and especially no obligations *of justice*, to future people. We may happen to care about them, we may choose to take their interests into account, but we owe them nothing. Planting a time bomb violates no obligations.[16]

One example is the social-contract theory. We cannot bargain, negotiate, or cooperate with those who will live long after us – and so a *contract* with distant future people

seems incoherent. However, while some social-contract theorists bite the bullet and reject intergenerational justice, most do not. Instead, they draw on a wide range of theoretical devices to accommodate intergenerational justice: assumptions about the motivations of present people; contracts between overlapping generations that somehow reach indefinitely into the future; the appointment of trustees or ombudsmen for the future; and imaginary intergenerational bargaining situations where the parties know neither when nor whether they exist.[17]

All intergenerational social contracts are controversial. They can also seem troublingly ad hoc. Suppose you really did believe that justice *is* cooperation for mutual advantage. Once you realize that future people cannot interact with us, wouldn't you simply lose interest in this contradictory non-topic called "intergenerational justice"? Won't any consistent social-contract theorist simply ignore intergenerational justice entirely? (Or, at best, treat it as an optional extension of the basic theory of justice between contemporaries?) Conversely, utilitarians argue that the fact that we *do* have obligations to distant future people demonstrates that morality is *not* ultimately about reciprocity at all.

Misplaced optimism

Obligations to future people raise problems of non-identity and reciprocity. Yet these puzzles have only recently become the focus of philosophical debate. The explanation is that, until very recently, moral philosophers concentrated exclusively on interactions between contemporaries. Future generations were only ever an afterthought. Philosophers saw no practical conflict between the interests of present and future people. If we do what is best for *present* people, then *future* people will inherit our stable liberal democratic institutions, thriving economy, and scientific advances. If we pursue our own interests – follow our natural inclinations – then future people will inevitably be better off than us. They will be healthier, wealthier, and enjoy longer and richer lives. What is good for us is also good for them. (This optimism is especially pronounced in social-contract theorists, notably Rawls.)[18]

Since the early 1970s, debates over ozone depletion, carbon emissions, greenhouse gases, and climate change have led philosophers to question this optimism. We now recognize the real threat of conflicts between the interests of present and future people. We also recognize that future people may be worse off than us. It may already be too late to prevent this; and any feasible proposal to minimize the harmful effects of climate change experienced by future people will involve a major reduction in the standard of living of affluent people living in developed countries today.

If conflicts between generations were rare, then the lack of intergenerational reciprocity could be dismissed as merely academic, rather than something that seriously undermines a moral theory's practical utility. (After all, every moral theory has some problems.) The possibility of real-world conflicts of interest between present and future

people gives the question of what we owe to future people a new urgency. Utilitarians argue that only they can offer coherent guidance here – and that the theoretical deficiencies of non-utilitarian theories can no longer be ignored.

Puzzles of aggregation

The primary focus of the vast literature on utilitarianism and future people is *aggregation*. Under the no difference view, different people choices per se present no new ethical issues. Unfortunately, such choices are not all alike. As Parfit notes, we can further divide different people choices into *same number* (where our choice affects who exists, but not how many people exist), and *different number* (where our choice affects how many people exist).[19] Different number choices raise many new difficulties. Suppose you could create any possible world, with any possible population. Which world should you choose? Because they base morality on the pursuit of the best possible consequences, utilitarians must answer this question. Utilitarians need a theory of aggregation – taking us from the values of individual lives to the values of possible populations. (Utilitarian theories of aggregation remain neutral between competing accounts of well-being. We will question this neutrality in the section on well-being, but we follow it for the time being.)

The utilitarian tradition offers two main accounts of aggregation. On the total view, the best outcome is the one that contains the greatest total amount of happiness. On the average view, the best outcome is the one that contains the highest average level of happiness. Classical utilitarians did not always clearly distinguish these views. This is understandable, as the two views must coincide in same number choices, where whatever maximizes the total also maximizes the average. But the two views often come apart in different number choices. Consider a choice between one possible future where a large population enjoys moderate happiness, and another where fewer people are very much happier. Suppose the first future has greater total happiness, while the second has higher average happiness. Which future is better in terms of human happiness?

The total view and the repugnant conclusion

The total view is the simplest theory of aggregation, and it has been the most popular view among utilitarian philosophers. (Economists, by contrast, have often favored the average view.) The basic argument for this view is simple. If we value happiness, then presumably we should aim to produce as much happiness as possible. The most famous objection to the total view dates back to Henry Sidgwick, and takes its modern name from Parfit.[20]

> *The repugnant conclusion*. For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable

population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living.

To see why the total view implies the repugnant conclusion, begin with a world (A) where ten billion people all have extremely good lives. Imagine a second world (B), with more than twice as many people, each of whom is only half as happy as the people in A. Total happiness in B exceeds that in A. Now repeat this process until we reach a world (Z) where a vast population have lives barely worth living. As each step increases total happiness, Z must be better than A.

Parfit finds this result "intrinsically repugnant."[21] If the total view yields this conclusion, then it is unacceptable. The repugnant conclusion is a classic example of a thought experiment that allegedly constitutes a decisive counterexample to a philosophical view. It is one of the organizing problems of contemporary intergenerational ethics.[22] Most philosophers begin their discussions by saying how they will deal with it – rejecting either Parfit's intuition that A is better than Z or the total view itself.

It is tempting to reject intuitions altogether. What matters is whether a conclusion follows from well-established premises, not whether it "appears" repugnant. But then what, other than some basic moral intuitions, could ground those premises? When philosophers say that they reject intuitions, this usually means that they reject some intuitions in favor of others. Non-utilitarians have the option of rejecting all intuitions about the comparative values of possible futures. They can refuse to say whether A is better or worse than Z. But utilitarians cannot take this option. Utilitarians need a theory of aggregation.

Some utilitarians reject intuitions regarding very large numbers. John Broome argues that "we have no reason to trust anyone's intuitions about very large numbers, however excellent their philosophy. Even the best philosophers cannot get an intuitive grasp of, say, tens of billions of people."[23] We should not abandon moral intuitions altogether, but instead build a theory on a foundation of everyday intuitions. Broome argues that those intuitions support the total view.

Others defend the total view by questioning Parfit's specific intuition about A and Z. Yew-Kwang Ng objects that our intuitions are guilty of "misplaced partiality."[24] We picture the A-lives as similar to our own, and imagine the A-people choosing between A and Z. If we were more impartial, we might see that Z is better than A because it contains more happiness. Along similar lines, other total utilitarians urge us to examine Z more closely. On the total view, we should create an extra life whenever doing so would raise the total happiness – whenever the extra life itself is worth living. If we imagine a numerical scale of well-being, then the lives in Z must be above zero. They are, by definition, "barely worth living." What would such lives be like? The phrase 'barely worth living' can evoke a life of frustration and pain: one that we would rather not live at

all. But a life we would rather not live is not worth living at all. If the Z-lives are like that, then the total view must hold that Z is worse than A. Indeed, Z would be even worse than an empty world where no one exists. Parfit's Z must contain *more* happiness than a world of ten billion flourishing lives. So it must be very different.

The average view

While some utilitarians defend the total view, others seek alternatives. The average view easily avoids the repugnant conclusion, as A has a higher average happiness than Z. However, the average view faces objections of its own. Many are variations of the *hermit problem*. Suppose everyone in the cosmos is extremely happy. We create a new person on a distant uninhabited planet. His life, while very good, is slightly below the cosmic average. Under the average view, we have made things worse; and whether we ought to have created the hermit in the first place depends on the happiness of people in distant corners of the cosmos, with whom our hermit will never interact. Both claims seem implausible. As Parfit puts it, the "mere addition" of lives worth living cannot make things worse, and our moral decisions should not depend on how happy the ancient Egyptians were.[25]

The hermit problem plays a similar dialectical role to the repugnant conclusion. Defenders of the average view have the same broad options. They can defend this conclusion, or deny that their theory implies it. One popular response is to limit our calculation of average happiness to those affected by our actions – thus removing the need to consider the welfare of distant people. But this still leaves Parfit's mere addition objection. Any proponent of the average view must bite the bullet and agree that the addition of any person with below-average happiness does make things worse.

The lexical view

Another popular account of aggregation is the lexical view.[26] Suppose you enjoy both Mozart and Muzak. Someone offers you a choice between one day of Mozart and as much Muzak as you like. If you opt for the former, saying that no amount of Muzak could match the smallest amount of Mozart, then you believe that Mozart is *lexically superior* to Muzak. The lexical view holds that some possible human lives are lexically superior to others.

The lexical view can avoid the repugnant conclusion. Suppose the creatures in A and Z belong to different species. Perhaps A contains flourishing human beings while Z is full of slugs. A is better, because ten billion human lives are more valuable than any number of slug lives. More controversially, a lexical view could also hold that ten billion flourishing human lives trump any number of human lives that are barely worth living – so that A can be better than Z even if both contain only human beings. (The lexical view also easily avoids the mere addition problem. Adding an extra life that is worth living

does not make things *worse*, even if that life is below the lexical level.)

Some reject the lexical view because it tells us to favor those who enjoy Mozart over Muzak-lovers. After all, Parfit suggests that we fall below the lexical level once the best things in life disappear. In practice, however, the lexical view might be extremely *egalitarian*. We could identify the lexical level, not with the best possible particular experiences or accomplishments, but with certain central human capacities, such as autonomy, the ability to pursue valuable goals, etc., so that any reasonably successful autonomous life is above the lexical level. In a society (such as our own) where a few people fall below the lexical level while most people are above it, this lexical view *now* gives priority to the worst-off people. It is better to raise one person above the lexical level than to benefit those *already* above it. In this context, a lexical view can be more egalitarian than either the total view or the average view.

Another problem for any lexical view is Parfit's *continuum objection*.

> Mozart and Muzak . . . seem to be in quite different categories. But there is a fairly smooth continuum between these two. Though Haydn is not as good as Mozart, he is very good. And there is other music which is not far below Haydn's, other music not far below this, and so on. Similar claims apply to the . . . other things which give most to the value of life . . . Since this is so, it may be hard to defend the view that what is best has more value than any amount of what is nearly as good.[27]

The lexical view must tell us where to draw the line – and why. How do we decide which possible *human* lives are above the lexical threshold, and which are below? Because the practical implications of the lexical view depend very largely on where we set the threshold, these are very significant decisions.

Discounting

Another debate in the literature on aggregation concerns the practice of *discounting* future harms and benefits. This discounting is relatively uncontroversial as a proxy for uncertainty, and to accommodate the remote possibility that there will be no future people. (Humanity might be wiped out by an asteroid strike, for instance.) There are also sound utilitarian reasons to discount if you are confident that future people will be richer than present people, or that technological advances will leave them better able to exploit any valuable resource. (Of course, this argument must be reversed if we expect future people to be worse off.) The controversial question is whether we should apply a *pure time preference*, which would entail that future happiness counts for less simply *because* it lies in the future. One common justification is that this pure time preference mirrors actual behavior. We do discount future benefits both to ourselves and to others.

Discount rates have a huge practical impact. Climate change provides a striking illustration. One prominent skeptical argument holds that the *future* benefit of preventing

climate change is not worth the *present* cost. Present funds do more good if devoted to the alleviation of present poverty. This cost–benefit analysis needs a social discount rate. Even a modest discount of 5 percent per annum makes it "uneconomic" to spend even one dollar today to avert a global catastrophe in five hundred years' time. (To be worth a dollar today, the catastrophe has to cost $137,466,652,006 at that future date.)[28] Different economists reach radically different conclusions on the basis of their divergent discount rates.[29]

The pure time preference is controversial among economists. In contrast, most utilitarian *philosophers* reject it and embrace temporal impartiality.[30] But, as we shall soon see, given the vast number of future people, a principled refusal to discount *at all* threatens to overwhelm all other moral concerns.

Infinite utility

Another set of aggregative puzzles arises from the possibility that the universe contains an infinite number of sentient beings.[31] If every possible future contains infinite value, then standard transfinite arithmetic implies that no possible future contains more value than any other. Therefore, no action is better or worse, in utilitarian terms, than any other. Suppose we begin with an infinite population where everyone has one unit of happiness, and then we give one person an additional unit. This does not increase total happiness. Indeed, even if we doubled *everyone's* happiness in our infinite population, total happiness would still remain the same.

Paradoxes of infinity threaten to paralyze utilitarianism. Such paradoxes arise if the universe is infinitely extended in time, or if it is spatially infinite and already contains an infinite number of inhabitants. Even a small probability of an infinite population is sufficient to give every action the same (infinite) expected value.

Some utilitarians avoid infinite utility by stipulation – limiting their attention to finite populations. This makes sense if our actions only affect our own planet (or galaxy), and if that planet (or galaxy) can only contain finitely many sentient beings. (This solution is especially congenial to those average utilitarians who restrict their attention to the well-being of agents in some local region.)

However, even if it cannot arise in practice, infinite utility is still a theoretical problem. A complete theory of aggregation must offer some principled way to determine when one possible infinite future is better or worse than another. One promising suggestion comes from Peter Vallentyne and Shelly Kagan.[32] Suppose the goodness of outcomes is based on some aggregation of local goodness – where possible locations for goodness include people, states of nature, and spatiotemporal regions. Vallentyne and Kagan then offer a principle applicable to all types of location: If w1 and w2 have exactly the same locations, and if, relative to any finite set of locations, w1 is better than w2, then w1 is better than w2. For example, if we double everyone's happiness, then we make things better. (The

locations are people's lives, and the result is better for every person.) Vallentyne and Kagan then extend this basic idea to other cases.

The Mere addition paradox

While the philosophical literature contains many other theories of aggregation, they all face similar problems to the total, average, and lexical theories. One focus of debate is Parfit's *mere addition paradox*.[33]

Parfit first imagines three possible futures:

- A contains 10 billion people with very flourishing lives.
- B contains 20 billion people with fairly flourishing lives. Each B-person is more than half as happy as each A-person.
- A+ contains 20 billion people split into two groups of 10 billion. One group is as happy as the A-people. The second group is less happy than the B-people, but their lives are still well worth living. Average happiness is lower in A+ than in B.

Parfit next compares each pair of possible futures:

- *B is not worse than A+*, because B has greater total happiness, greater average happiness, and less inequality.
- *A+ is not worse than A*, because the only difference is that A+ contains 10 billion additional happy lives, and the mere addition of worthwhile lives cannot make an outcome worse.
- *B is worse than A*, because if we deny the repugnant conclusion, then there must be some point when the shift from A to B makes things worse.

But this triad of comparative judgments is contradictory. If B is not worse than A+, and A+ is not worse than A, then how can B be worse than A? Parfit argues that we cannot avoid the repugnant conclusion and at the same time claim that the mere addition of happy lives never makes things worse.

Parfit's paradox has generated a host of other "impossibility results," each arguing that some set of intuitive requirements on a theory of aggregation cannot all be satisfied.[34]

Some utilitarians argue that our judgments of the comparative values of different possible futures are intransitive.[35] That is, we cannot infer that A is better than C just because A is better than B and B is better than C. Therefore, Parfit's three pairwise comparisons are not strictly contradictory, and the real lesson is that no single theory of aggregation covers all possible pairwise comparisons. Against this, Broome argues that 'better than' is an intrinsically transitive notion, and we can only dissolve the mere addition paradox by embracing the repugnant conclusion.[36]

Another way to avoid the mere addition paradox is a relativized model of value, where we evaluate different possible worlds relative to the interests of the people who live in them. Perhaps A is better than B from one perspective, while B is better from another.[37] For instance, Melinda Roberts defends a *person-affecting consequentialism* where, instead of maximizing total happiness, we aim to maximize the happiness of each individual by giving her the best life she could possibly have enjoyed.[38]

Puzzles of right action

We now turn to a series of puzzles that arise for any utilitarian intergenerational ethic, whatever its account of aggregation. Any utilitarian moral theory complements its theory of value with a theory of right action – telling us how to promote the good. The simplest theory of right action is act consequentialism: the right act in any situation is whatever produces the best consequences. Most recent utilitarian discussions of aggregation implicitly assume act consequentialism. However, in combination with any of the theories of aggregation explored in the previous section, act consequentialism yields several very counterintuitive results.[39]

First, act consequentialism makes our obligations to future people extremely demanding.[40] Given the enormous number of future people, and the significance of our potential impact on them, intergenerational ethics will swamp all other ethical considerations. Future people thus provide a very striking case of a perennial problem for utilitarianism: its demandingness. The most-discussed example concerns our obligations to people currently living in poverty in distant lands. As there are very many such people whom I could assist, my utilitarian obligations will swamp all my personal projects and special obligations. And needy *future* people vastly outnumber needy present people. Therefore, utilitarianism may be even more demanding in relation to future people than in the case of impoverished present people.

Second, because it requires agents to always perform whatever action has the best consequences, act consequentialism severely curtails reproductive freedom. Every moral agent must create new people whenever those people would have happy lives (or, in the case of the average view, whenever their happiness would be above average). (Alternatively, if I could do more good by devoting my resources to the needs of already existing people, then I may be *prohibited* from having any children at all. But this merely replaces one curtailment of reproductive freedom with another.)

A third objection is that act consequentialism denies a compelling commonsense *procreation asymmetry* which holds that, while there is a moral obligation not to create a person whose life would be *not* worth living, there is no parallel obligation to create a person whose life would be worth living. So long as we seek only to maximize total (or average) happiness, we cannot recognize this asymmetry.

All three objections arise for *any* account of aggregation. So long as we seek to maximize impersonal value, however that notion is cashed out, we must accept a very demanding moral theory that obliterates reproductive freedom and cannot recognize the procreation asymmetry.

For each objection, an extreme act consequentialist could bite the bullet. If morality is about the promotion of impersonal value, and if we retain a commitment to temporal impartiality, then, given the present state of the world, we must accept an extremely demanding morality and simply reject the asymmetries of commonsense morality.

Utilitarians looking for precedents here might appeal to J. S. Mill's skepticism about reproductive freedom in chapter 5 of *On Liberty*.[41] However, Mill himself certainly never embraced the extremism of contemporary act consequentialism. And most utilitarians seek a more moderate intergenerational ethic. One obvious strategy is to begin with standard responses to the demandingness objection, as this is a more general problem. Act consequentialism is already very demanding without taking account of future people, and utilitarians have developed many responses to the demandingness problem. Perhaps some existing response can also help us with future people.

We might begin by reexamining our intuitive reactions to Parfit's thought experiments. Many people have stronger intuitions regarding our obligations to future people than they do about the comparative values of possible futures. Rejection of the repugnant conclusion might be motivated, not by the abstract claim that Z is worse than A, but by the insistence that the inhabitants of A are not obliged to transform their world into Z. The act consequentialist cannot separate these two thoughts. Anyone who has a choice between two possible futures *must* opt for the better one. Moderate utilitarians might offer a more nuanced response. They can deny that anyone has an obligation to turn an A-world into a Z-world, without having to argue that A is a better outcome than Z.

One simple option is to reject temporal impartiality and institute a pure time preference *in practice*. Although the welfare of future people is as valuable as ours, perhaps we are entitled to discount it in our moral deliberations. But this move faces the objection that *mere* distance in time has no more moral significance than mere distance in space. A more nuanced approach, drawing on recent work by Samuel Scheffler, would be to give each agent permission to give special weight to her own interests (and perhaps those of her nearest and dearest) – against both those in the future and those currently living in distant lands.[42]

A more systematic option is to combine our preferred theory of aggregation not with act consequentialism but with rule consequentialism, where the right action is whatever follows from the set of rules whose internalization by everyone would produce the best consequences.[43] Unlike act consequentialism, rule consequentialism can respect the intuition behind the rejection of the repugnant conclusion without rejecting the total view. Because it endorses many familiar non-utilitarian moral distinctions, rule

consequentialism can also accommodate features of commonsense morality such as reproductive freedom and the procreation asymmetry.[44]

Indeed, rule consequentialism can also borrow elements from alternative theories of aggregation – such as lexical thresholds, the focus on average well-being, the restriction to finite populations, and discounting – and incorporate them into its theory of right action. The ideal code of rules – the one that best promotes total happiness – may encourage individual agents to depart from the total view in their moral deliberations (just as it encourages them to depart from act consequentialism). For instance, the lexical view might be interpreted as capturing a *practical obligation* to raise everyone in a given society above a certain threshold. Thus interpreted, the lexical threshold is context-dependent. Its precise location will differ from one deliberative context to another. Or a society might set itself the policy goal of maximizing the average quality of life of its grandchildren. Such moves enable us to retain all the theoretical advantages of the total view as an account of the comparative values of outcomes, while also respecting those intuitions that reject any obligation to maximize total value.[45]

Rethinking well-being

Like other moral philosophers, utilitarians typically treat obligations to future people as an extension of a basic moral theory designed for relations among contemporaries. We first select our preferred accounts of right action, human well-being, and aggregation, and then apply these to the intergenerational case. This priority is questionable, however, as reflection on future people may impact on general debates within utilitarian theory. The previous section explores the impact on utilitarian accounts of right action. We now examine the impact on theories of well-being.

We noted earlier that utilitarian discussions of future people seek to remain neutral between competing theories of well-being, as evidenced by the widespread adoption of "place-holders" such as Parfit's phrase 'whatever makes life worth living'.[46] However, some accounts of well-being extend more naturally to future people than others. In particular, both preference utilitarianism and hedonism are inadequate when applied to distant future people, as they can find no fault with a decision to create future beings who lack some specific good but have no preference for what they have lost and would find no pleasure in it. (Suppose we destroy a beautiful natural environment, but then ensure that future people have no appreciation for natural beauty.) Even Peter Singer, who has been the most prominent contemporary defender of preference utilitarianism, has recently conceded that we need objective goods to make sense of our real obligations to future people.[47]

Debates over aggregation are also not neutral between competing accounts of well-being. The total and average views both work best with theories of well-being that offer a numerical scale, such as hedonism or preference theory. By contrast, the lexical view is

most naturally combined with an objective list account, as any attempt to map a lexical threshold onto a continuous scale of total pleasure or preference-satisfaction is bound to seem implausibly ad hoc in the face of Parfit's continuum objection.

The mere fact that our utilitarian calculations must include the well-being of future people thus has an impact on utilitarian theory. That impact is potentially much deeper once we factor in the threat of a broken future.

Utilitarianism for a broken world

The prospect of a broken future, where future people are worse off than present people and many optimistic background assumptions of contemporary moral and political philosophy no longer hold, may seem of limited *theoretical* interest. Surely utilitarians should simply apply the best utilitarian theory to changing circumstances? Unfortunately, things are not so straightforward. Declining well-being may impact on utilitarianism in a wide variety of ways. The rest of this section sketches some possibilities.[48]

We begin with the question of aggregation. We saw earlier, in the section on the average view, that that view has especially counterintuitive implications regarding the addition of below-average lives. If well-being increases over time, then most future lives will be above average, so these difficulties do not arise in regard to future people. By contrast, if we face a broken future where the vast majority of future lives will be below average, then the average view implies that humanity should have no future.

Similarly, we saw earlier that the lexical view is plausible *if* we can safely assume that most future people can live above the lexical threshold. However, if that threshold is defined with reference to current standards of living that future people cannot hope to enjoy, then it may be inevitable that future people will fall below our lexical threshold. In this situation, a lexical view is likely to undervalue the prospects of future people.

A broken future thus increases the comparative appeal of the total view as against either the average view or the lexical view. However, the total view is especially demanding in a broken world. An alternative is to relativize the relevant standard, whether average well-being or lexical threshold, so that it varies from one generation to the next. Such relativization is problematic, and perhaps incoherent, if we are developing an objective account of aggregation. But it makes sense if (following the suggestion of the section on right action above) we think of lexical thresholds (or desired levels of average well-being) not as components of our account of aggregation, but as independently motivated elements of our theory of right action.

A broken future increases the demands of utilitarianism, especially on comparatively affluent present people. At first glance, this seems to count against act utilitarianism, as it is already the most demanding theory of right action. However, act utilitarians are already reconciled to an extremely (and counterintuitively) demanding moral theory – and to the

uncomfortable thought that changing global circumstances might significantly increase those demands. A broken future may thus be more troubling for moderate utilitarians, notably rule utilitarians. If great sacrifices by present people would significantly alleviate the burdens faced by countless generations of worse-off future people – perhaps by reducing the impact of anthropogenic climate change – then it is very unlikely that the best utilitarian code of rules we could teach to the next generation will make only moderate demands of *us*. A broken future may thus deprive rule utilitarianism of one of its greatest advantages over act utilitarianism – its ability to make only moderate demands.

Similar problems face other familiar attempts to reconcile utilitarianism with prima facie non-utilitarian moral ideas – such as rights, liberty, or democracy. Any utilitarian defense of these elements is contingent, and thus vulnerable to empirical change. Is democracy the best utilitarian solution to the challenges posed by climate change? Can present freedoms of speech, action, or reproduction trump the needs and survival of future people? Can utilitarians, or anyone else, make sense of rights in a world whose resources are not sufficient for everyone to survive?

Utilitarians may respond that, because so much of our conventional moral thought presupposes optimism about the future, we cannot take the broken future seriously without abandoning many cherished moral certainties. Our moral intuitions, tailored to a disappearing affluent world, are no longer a reliable guide. Utilitarianism is often attacked for its willingness to think the unthinkable. The English Roman Catholic philosopher Elizabeth Anscombe went so far as to describe utilitarian thinking as the product of a corrupt mind.[49] Perhaps, in a broken world, where the unthinkable must be thought, this willingness becomes, not a vice, but a necessary virtue.

Notes

1. While it is often attributed to Bentham, this precise phrase is apparently not found in any of his extant writings. The attribution goes back to Mill: J. S. Mill, *Utilitarianism, Collected Works*, vol. x, p. 257.

2. See, e.g., Mulgan, *Ethics for a Broken World*; and Mulgan, "Utilitarianism for a Broken World."

3. Parfit, *Reasons and Persons*, p. 356.

4. Parfit, *Reasons and Persons*, p. 367.

5. Parfit, *Reasons and Persons*, p. 358 and p. 371.

6. Parfit, *Reasons and Persons*, p. 362.

7. See the essays in Roberts and Wasserman, *Harming Future Persons*.

8. Mulgan, *Future People*, chapter 2. For discussions of a variety of contractualist solutions, see the essays in Gosseries and Meyer, *Intergenerational Justice*.

9. Heyd explores this approach throughout his book *Genethics*.

10. See the essays in Roberts and Wasserman, *Harming Future Persons*.

11. To further complicate matters, some utilitarians have developed person-affecting versions of the theory – explicitly designed to reject Parfit's no difference view. (Roberts, "A New Way of Doing the Best That We Can.")

12. Kumar, "Who Can Be Wronged?", p. 111.

13. Kumar, "Who Can Be Wronged?", p. 112.

14. Parfit, *On What Matters*, chapter 22.

15. Parfit, *On What Matters*, chapter 22.

16. Explicit defenses of time bombs are rare. However, such defenses are implicit in many non-utilitarian discussions.

17. See, again, the essays in Gosseries and Meyer, *Intergenerational Justice*.

18. Rawls, *A Theory of Justice*, pp. 251–259.

19. Parfit, *Reasons and Persons*, p. 356.

20. Sidgwick, *The Methods of Ethics*, pp. 415–416 (in book IV, chapter 1, section 2); Parfit, *Reasons and Persons*, p. 388.

21. Parfit, *Reasons and Persons*, p. 390.

22. See, e.g., the essays in Ryberg and Tännsjö, *The Repugnant Conclusion*.

23. Broome, *Weighing Lives*, pp. 57–58.

24. Ng, "What Should We Do about Future Generations?", p. 242.

25. Parfit, *Reasons and Persons*, p. 420.

26. Crisp, "Utilitarianism and the Life of Virtue"; Parfit, "Overpopulation and the Quality of Life"; and Mulgan, *Future People*, chapter 3.

27. Parfit, "Overpopulation and the Quality of Life," p. 164.

28. With a discount rate of 5 percent, the present value of a single dollar n years from now is $(0.95)^n$. Therefore, the present value of a single dollar in 500 years' time is $(0.95)^{500}$.

29. For economic discussion, especially in relation to climate change, see Stern, *Stern Review*, and Nordhaus, *The Challenge of Global Warming*, pp. 143–161.

30. The classic philosophical discussion is Cowen and Parfit, "Against the Social Discount Rate."

31. See, e.g., Vallentyne, "Utilitarianism and Infinite Utility."

32. Vallentyne and Kagan, "Infinite Value," p. 9.

33. Parfit, *Reasons and Persons*, pp. 419–442. For responses to the paradox, see, e.g., Carlson, "Mere Addition and Two Trilemmas"; and Temkin, "Intransitivity and the Mere Addition Paradox."

34. See, e.g., Arrhenius, "An Impossibility Theorem for Welfarist Axiologies."

35. Temkin, "Intransitivity and the Mere Addition Paradox"; and Rachels, "A Set of Solutions to Parfit's Problems."

36. Broome, *Weighing Lives*, pp. 50–63.

37. Dasgupta, "Savings and Fertility," pp. 120–125.

38. Roberts, "A New Way of Doing the Best That We Can."

39. This section draws on Mulgan, *Future People*, chapter 3.

40. Recent discussions of the demands of utilitarianism include Cullity, *The Morality of Affluence*; Mulgan, *The Demands of Consequentialism*; and Murphy, *Moral Demands in a Non-Ideal World*.

41. J. S. Mill, *On Liberty, Collected Works*, vol. XVIII, pp. 301–302.

42. Scheffler, *The Rejection of Consequentialism*, p. 4; and Mulgan, *Future People*, chapter 4.

43. Hooker, *Ideal Code, Real World*, p. 32.

44. Mulgan, *Future People*, chapter 6.

45. Mulgan, *Future People*, p. 174.

46. Parfit, *Reasons and Persons*, p. 387.

47. Singer, *Practical Ethics*, p. x; and Mulgan, "What Is Good for the Distant Future?"

48. This section draws on Mulgan, *Ethics for a Broken World*, part 2; and Mulgan, "Utilitarianism for a Broken World."

49. Anscombe, “Modern Moral Philosophy,” pp. 16–17.

Bibliography

Adams, R. M. *Finite and Infinite Goods: A Framework for Ethics*. New York: Oxford University Press, 1999.

Adams, R. M. "Motive Utilitarianism," *Journal of Philosophy* 73 (1976), 467–481.

Anomaly, J. "Nietzsche's Critique of Utilitarianism," *Journal of Nietzsche Studies* 29 (2005), 1–15.

Anscombe, G. E. M. "Modern Moral Philosophy," *Philosophy* 33 (1958), 1–19.

Aristotle *Nicomachean Ethics*, ed. and trans. R. Crisp, Cambridge: Cambridge University Press, 2000.

Aristotle *Politics*, trans. B. Jowett, in *Aristotle, The Complete Works of Aristotle*, 2 vols., ed. J. Barnes, vol. II, pp. 1986–2129.

Armitage, D. "Globalizing Jeremy Bentham," *History of Political Thought* 32 (2011), 63–82.

Arneson, R. J. "Benthamite Utilitarianism and Hard Times," *Philosophy and Literature* 2 (1978), 60–75.

Arneson, R. J. "Human Flourishing versus Desire Satisfaction," *Social Philosophy and Policy* 16 (1999), 113–142.

Arneson, R. J. "Sophisticated Rule Consequentialism: Some Simple Objections," *Philosophical Issues* 15 (2005), 235–251.

Arrhenius, G. "An Impossibility Theorem for Welfarist Axiologies," *Economics and Philosophy* 16 (2000), 247–266.

Austin, J. *The Province of Jurisprudence Determined*, ed. W. E. Rumble, Cambridge: Cambridge University Press, 1995.

Avila-Martel, A. de "The Influence of Bentham on the Teaching of Penal Law in Chile," *Bentham Newsletter* 5 (1981), 22–28.

Aydelotte, W. O. "The England of Marx and Mill as Reflected in Fiction," *Journal of Economic History* 8 supplement (1948), 42–58.

Bales, R. E. "Act-Utilitarianism: Account of Right-Making Characteristics or Decision-Making Procedure?" *American Philosophical Quarterly* 8 (1971), 257–265.

Balguy, J. *The Foundation of Moral Goodness: or A Further Inquiry into the Original*

of Our Idea of Virtue, London: J. Pemberton: 1728.

Baron, M. *Kantian Ethics Almost without Apology*, Ithaca, NY: Cornell University Press, 1995.

Barry, B. *Theories of Justice*, Berkeley, CA: University of California Press, 1989.

Beccaria, C. *Of Crimes and Punishments [Dei delitti e delle pene]*, trans. Jane Grigson, New York: Marsilio, 1996.

Bentham, J. *A Comment on the Commentaries and A Fragment on Government*, eds. J. H. Burns and H. L. A. Hart, London: Athlone Press, 1977.

Bentham, J. Bentham Papers, University College London Library. References refer to box number and folio number.

Bentham, J. *Church-of-Englandism and Its Cathechism Examined*, eds. J. E. Crimmins and C. Fuller, Oxford: Clarendon Press, 2011.

Bentham, J. *Constitutional Code*, vol. I, eds. F. Rosen and J. H. Burns, Oxford: Clarendon Press, 1983.

Bentham, J. *The Correspondence of Jeremy Bentham*, vols. I–XIV, London: Athlone Press, 1968–81, and Oxford: Clarendon Press, 1984–in progress.

Bentham, J. *Deontology together with A Table of the Springs of Action and the Article on Utilitarianism*, ed. A. Goldworth, Oxford: Clarendon Press, 1983.

Bentham, J. *An Introduction to the Principles of Morals and Legislation*, eds. J. H. Burns and H. L. A. Hart, Oxford: Clarendon Press, 1996.

Bentham, J. *"Legislator of the World": Writings on Codification, Law, and Education*, eds. P. Schofield and J. Harris, Oxford: Clarendon Press, 1998.

Bentham, J. *Of the Limits of the Penal Branch of Jurisprudence*, ed. P. Schofield, Oxford: Clarendon Press, 2010.

Bentham, J. *Principles of International Law* in *The Works of Jeremy Bentham*, vol. II, ed. J. Bowring, Edinburgh: William Tait, 1843, pp. 535–560.

Bentham, J. *Rationale of Judicial Evidence* in *The Works of Jeremy Bentham*, vols. VI–VII, ed. J. Bowring, Edinburgh: William Tait, 1843.

Bentham, J. *Traités de législation civile et pénale . . . Publiés en français par Ét. Dumont de Genève*, 3 tom., 2nd edn., Paris: Bossange, Père et Fils; Rey et Gravier, 1820. Eng. trans., *Theory of Legislation* . . . by R. Hildreth, 2 vols., 1840; 2nd edn., 1864, rept. with introduction by J. E. Crimmins, Bristol, UK: Thoemmes Continuum, 2004.

Bentham, J. *The Works of Jeremy Bentham*, ed. J. Bowring, 11 vols., Edinburgh: William Tait, 1838–43.

Berger, F. R. *Happiness, Justice, and Freedom: The Moral and Political Philosophy of John Stuart Mill*, Berkeley, CA: University of California Press, 1984.

Bergström, L. *The Alternatives and Consequences of Actions: An Essay on Certain Fundamental Notions in Teleological Ethics*, Stockholm: Almqvist & Wiksell, 1966.

Bergström, L. "Reflections on Consequentialism," *Theoria* 62 (1996), 74–94.

Berkeley, G. *Passive Obedience, Or the Christian Doctrine of not Resisting the Supreme Power, Proved and Vindicated upon the Principles of the Law of Nature*, Dublin: F. Dickson, 1712.

Berman, D. "The Jacobitism of Berkeley's Passive Obedience," *Journal of the History of Ideas* 47 (1986), 309–319.

Blackorby, C., W. Bossert, and D. Donaldson, *Population Issues in Social-Choice Theory, Welfare Economics and Ethics*, Cambridge: Cambridge University Press, 2005.

Blamires, C. *The French Revolution and the Creation of Benthamism*, London: Palgrave Macmillan, 2008.

Boonin, D. *The Problem of Punishment*, Cambridge: Cambridge University Press, 2008.

Bradford, W. *Of Plymouth Plantation: 1620–1647*, ed. S. E. Morison, New York: Alfred A. Knopf, 1952.

Bradley, B. "Against Satisficing Consequentialism," *Utilitas* 18 (2006), 97–108.

Bradley, B. *Well-Being and Death*, New York: Oxford University Press, 2009.

Bradley, F. H. *Ethical Studies*, London: H. S. King, 1876.

Brandt, R. B. *A Theory of the Good and the Right*, Oxford: Oxford University Press, 1979.

Brandt, R. B. *Ethical Theory: The Problems of Normative and Critical Ethics*, Englewood Cliffs, NJ: Prentice-Hall, 1959.

Brandt, R. B. *Facts, Values, and Morality*, Cambridge: Cambridge University Press, 1996.

Brandt, R. B. "Fairness to Indirect Optimific Theories in Ethics," in R. B. Brandt, *Morality, Utilitarianism, and Rights*, Cambridge: Cambridge University Press, 1992, pp. 137–157.

Brandt, R. B. "Some Merits of One Form of Rule-Utilitarianism," in R. B. Brandt, *Morality, Utilitarianism, and Rights*, Cambridge: Cambridge University Press, 1992, pp. 111–136.

Brandt, R. B. "Toward a Credible Form of Utilitarianism," in *Morality and the Language of Conduct*, eds. H.-N. Castañeda and G. Nakhnikian, Detroit: Wayne State University Press, 1965, pp. 107–143.

Brandt, R. B. "Utilitarianism and the Laws of War," *Philosophy and Public Affairs* 1 (1972), 145–165.

Broad, C. D. *Five Types of Ethical Theory*, London: K. Paul, Trench, Trubner, 1930. See especially chapter VI, "Sidgwick."

Broome, J. "A Reply to Sen," *Economics and Philosophy* 7 (1991), 285–287.

Broome, J. "'Utility'," *Economics and Philosophy* 7 (1991), 1–12.

Broome, J. *Weighing Goods: Equality, Uncertainty and Time*, Oxford: Basil Blackwell, 1991.

Broome, J. *Weighing Lives*, Oxford: Oxford University Press, 2004.

Brown, C. "Blameless Wrongdoing and Agglomeration: A Response to Streumer," *Utilitas* 17 (2005), 222–225.

Brown, J. *Essays on the Characteristics*, London: Davis, 1751.

Burch-Brown, J. M. "Clues for Consequentialists," forthcoming in *Utilitas* 26 (2014).

Butler, J. *The Works of Joseph Butler*, ed. W. E. Gladstone, Oxford: Clarendon Press, 1896.

Bykvist, K. "How to Do Wrong Knowingly and Get Away with It," in *Neither/Nor: Philosophical Papers Dedicated to Erik Carlson on the Occasion of His Fiftieth Birthday*, eds. E. Carlson *et al.*, Uppsala: Department of Philosophy, Uppsala University, Uppsala Philosophical Studies 58 (2011).

Bykvist, K. *Utilitarianism: A Guide for the Perplexed*, London: Continuum, 2010.

Byron, M. (ed.) *Satisficing and Maximizing: Moral Theorists on Practical Reason*, Cambridge: Cambridge University Press, 2004.

Cameron, F. *Nietzsche and the 'Problem' of Morality*, New York: Peter Lang, 2002.

Caplan, B. "Eureka! Economic Illiteracy as Mental Substitution," in Library of Economics and Liberty, Liberty Fund, Inc., at http://econlog.econlib.org/archives/2012/01/eureka_economic.html, 2012.

Carlson, E. *Consequentialism Reconsidered*, Dordrecht: Kluwer Academic Publishers, 1995.

Carlson, E. "Mere Addition and Two Trilemmas of Population Ethics," *Economics and Philosophy* 14 (1998), 283–306.

Carlyle, T. *On Heroes, Hero-Worship, and the Heroic in History*, eds. M. K. Goldberg, J. J. Brattin, and M. Engel, Berkeley, CA: University of California Press, 1993.

Carmichael, G. *Natural Rights on the Threshold of the Scottish Enlightenment: The Writings of Gershom Carmichael*, eds. J. Moore and M. Silverthorne, trans. M. Silverthorne, Indianapolis: Liberty Fund, 2002.

Carson, T. L. "Church-of-Englandism and Its Catechism examined . . . by Jeremy Bentham," *Quarterly Review* 21 (1819), 167–177.

Carson, T. L. *Value and the Good Life*, Notre Dame, IN: University of Notre Dame Press, 2000.

Cocking, D. and J. Oakley, "Indirect Consequentialism, Friendship, and the Problem of Alienation," *Ethics* 106 (1995), 86–111.

Cohen, G. A. *Self-Ownership, Freedom, and Equality*, Cambridge: Cambridge University Press, 1995.

Cohen, R. "Can You Forgive Him? How a Book and a Friendship Went up in Flames," *New Yorker*, November 8, 2004, pp. 48–65.

Collini, S., D. Winch, and J. Burrow, *That Noble Science of Politics: A Study in Nineteenth-Century Intellectual History*, Cambridge: Cambridge University Press, 1983.

Conway, S. "Bentham on Peace and War," *Utilitas* 1 (1989), 82–101.

Cooper, T. *Lectures on the Elements of Political Economy*, 2nd edn., New York: A. M. Kelley, 1971.

Cooper, T. *Philosophical Writings of Thomas Cooper*, 3 vols., ed. U. Thiel, Bristol, UK: Thoemmes Press, 2001.

Cooper, T. "Slavery," *Southern Literary Journal and Magazine of Arts* 1 (1835), 188–193.

Cowen, T. and D. Parfit, "Against the Social Discount Rate," in *Justice Between Age Groups and Generations*, eds. P. Laslett and J. Fishkin, New Haven, CT: Yale University Press, 1992, pp. 144–161.

Crimmins, J. E. "Religion, Utility and Politics: Bentham versus Paley," in *Religion,*

Secularization and Political Thought: Thomas Hobbes to J. S. Mill, ed. J. E. Crimmins, London: Routledge, 1989, pp. 130–152.

Crimmins, J. E. *Secular Utilitarianism: Social Science and the Critique of Religion in the Thought of Jeremy Bentham*, Oxford: Clarendon Press, 1990.

Crimmins, J. E. *Utilitarian Philosophy and Politics: Bentham's Later Years*, London and New York: Continuum, 2011.

Crimmins, J. E. *Utilitarians and Religion*, Bristol: Thoemmes Press, 1998.

Crimmins, J. E. and M. G. Spencer (eds.) *Utilitarians and Their Critics in America*, 1789–1914, 4 vols., with an introduction by J. E. Crimmins, Bristol, UK: Thoemmes Continuum, 2005.

Crisp, R. "Hedonism Reconsidered," *Philosophy and Phenomenological Research* 73 (2006), 619–645.

Crisp, R. *Reasons and the Good*, New York: Oxford University Press, 2006.

Crisp, R. *Routledge Philosophy Guidebook to Mill on Utilitarianism*, London: Routledge, 1997.

Crisp, R. "Utilitarianism and the Life of Virtue," *Philosophical Quarterly* 42 (1992), 139–160.

Cumberland, R. *A Treatise of the Laws of Nature [De Legibus Naturae]*, trans. J. Maxwell, ed. J. Parkin, Indianapolis: Liberty Fund, 2005.

Cummiskey, D. *Kantian Consequentialism*, New York: Oxford University Press, 1996.

Dancy, J. *Moral Reasons*, Oxford: Blackwell Publishers, 1993.

Darwall, S. "Norm and Normativity," in *The Cambridge History of Eighteenth-Century Philosophy*, vol. II, ed. K. Haakonssen, Cambridge: Cambridge University Press, 2006, pp. 987–1025.

Darwall, S. "Valuing Activity," *Social Philosophy and Policy* 16 (1999), 176–196.

d'Aspremont, C. and L. Gevers, "Equity and the Informational Basis of Collective Choice," *Review of Economic Studies* 44 (1977), 199–209.

Dasgupta, P. "Savings and Fertility: Ethical Issues," *Philosophy and Public Affairs* 23 (1994), 99–127.

Dinwiddy, J. R. "Bentham and the Early Nineteenth Century," *Bentham Newsletter* 8 (1984), 15–33.

Diogenes Laërtius, *The Lives and Opinions of Eminent Philosophers*, trans. C. D.

Yonge, London: Henry G. Bohn, 1853.

Donagan, A. "Is There a Credible Form of Utilitarianism?" in *Contemporary Utilitarianism*, ed. M. Bayles, Gloucester MA: Peter Smith, 1978, pp. 187–202.

Donner, W. *The Liberal Self: John Stuart Mill's Moral and Political Philosophy*, Ithaca, NY: Cornell University Press, 1991.

Donner, W. "Mill's Moral and Political Philosophy," in W. Donner and R. Fumerton, *Mill*, Malden, MA: Wiley-Blackwell, 2009, pp. 13–143.

Donner, W. "Mill's Theory of Value," in *The Blackwell Guide to Mill's Utilitarianism*, ed. H. R. West, Malden, MA, and Oxford: Blackwell Publishing, 2006, pp. 117–138.

Dorsey, D. "Consequentialism, Metaphysical Realism, and the Argument from Cluelessness," *Philosophical Quarterly* 62 (2012), 48–70.

Donner, W. "Three Arguments for Perfectionism." *Noûs* 44 (2010), 59–79.

Driscoll, E. A. "The Influence of Gassendi on Locke's Hedonism," *International Philosophical Quarterly* 12 (1972), 87–110.

Driver, J. *Consequentialism*, Abingdon: Routledge, 2012.

Driver, J. *Uneasy Virtue*, Cambridge: Cambridge University Press, 2001.

Driver, J. "What the Objective Standard is Good For," in *Oxford Studies in Normative Ethics*, vol. II, ed. M. Timmons, Oxford: Oxford University Press, 2012, pp. 28–44.

Dworkin, R. *Justice for Hedgehogs*, Cambridge, MA: Harvard University Press, 2011.

Dworkin, R. "What Is Equality? Part 1: Equality of Welfare," *Philosophy and Public Affairs* 10 (1981), 185–246.

Eggleston, B. "Does Participation Matter? An Inconsistency in Parfit's Moral Mathematics," *Utilitas* 15 (2003), 92–105.

Eggleston, B. "Practical Equilibrium: A Way of Deciding What to Think about Morality," *Mind* 119 (2010), 549–584.

Eggleston, B. "Rejecting the Publicity Condition: The Inevitability of Esoteric Morality," *Philosophical Quarterly* 63 (2013), 29–57.

Ellis, B. "Retrospective and Prospective Utilitarianism," *Noûs* 15 (1981), 325–339.

Empson, W. "Bentham's Rationale of Evidence," *Edinburgh Review, or Critical Journal* 48 (1828), 457–520.

Engelmann, S. G. "Imagining Interest," *Utilitas* 13 (2001), 289–322.

Epicurus, *The Extant Remains*, trans. Cyril Bailey, Oxford: Clarendon Press, 1926.

Ewing, A. C. *The Definition of Good*, London: Routledge & Kegan Paul, 1947.

Feinberg, J. *Social Philosophy*, Englewood Cliffs, NJ: Prentice-Hall, 1973.

Feldman, F. "Actual Utility, the Objection from Impracticality, and the Move to Expected Utility," *Philosophical Studies* 129 (2006), 49–79.

Feldman, F. "Adjusting Utility for Justice," *Philosophy and Phenomenological Research* 55 (1995), 567–585.

Feldman, F. *Doing the Best We Can: An Essay in Informal Deontic Logic*, Dordrecht: D. Reidel, 1986.

Feldman, F. *Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism*, Oxford: Clarendon Press, 2004.

Feldman, F. "True and Useful: On the Structure of a Two Level Normative Theory," *Utilitas* 24 (2012), 151–171.

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

Feldman, F. "What We Learn from the Experience Machine," in *The Cambridge Companion to Nozick's Anarchy, State, and Utopia*, ed. R. M. Bader and J. Meadowcroft, Cambridge: Cambridge University Press, 2011, pp. 59–86.

Ferkany, M. "The Objectivity of Wellbeing," *Pacific Philosophical Quarterly* 93 (2012), 472–492.

Fleurbaey, M., B. Tungodden, and P. Vallentyne, "On the Possibility of Nonaggregative Priority for the Worst Off," *Social Philosophy and Policy* 26 (2009), 258–285.

Foot, P. "Utilitarianism and the Virtues," *Mind* 94 (1985), 196–209.

Foot, P. *Virtues and Vices and Other Essays in Moral Philosophy*, Berkeley, CA: University of California Press, 1978.

Forbes, D. "James Mill and India," *Cambridge Journal* 5 (1951), 19–33.

Force, P. "Helvétius as an Epicurean Political Theorist," in *Epicurus in the Enlightenment*, eds. N. Leddy and A. S. Lifschitz, Oxford: Voltaire Foundation, 2009, pp. 105–118.

Force, P. *Self-Interest before Adam Smith: A Genealogy of Economic Science*, Cambridge: Cambridge University Press, 2003.

Forero, J. "Leaving the wild, and rather liking the change," *New York Times*, May 11, 2006; at www.nytimes.com/2006/05/11/world/americas/11colombia.html.

Foucault, M. *Discipline and Punish: The Birth of the Prison*, New York: Pantheon, 1977.

Fuchs, A. E. "Mill's Theory of Morally Correct Action," in *The Blackwell Guide to Mill's Utilitarianism*, ed. H. R. West, Malden, MA, and Oxford: Blackwell, 2006, pp. 139–158.

Gaus, G. "On the Difficult Virtue of Minding One's Own Business: Towards the Political Rehabilitation of Ebenezer Scrooge," *The Philosopher: A Magazine for Free Spirits* 5 (1997), 24–28; at www.gaus.biz/scrooge.pdf.

Gauthier, D. "Rule-Utilitarianism and Randomization," *Analysis* 25 (1965): 68–69.

Gay, J. "Preliminary Dissertation Concerning the Fundamental Principle of Virtue or Morality," in W. King, *An Essay on the Origin of Evil*, vol. I, 2nd edn., ed. E. Law, London: 1732, pp. xxviii–lvii.

Gensler, H. J. "Paradoxes of Subjective Obligation," *Metaphilosophy* 18 (1987), 208–213.

Gert, B. *Morality: Its Nature and Justification*, revised edn., Oxford: Oxford University Press, 2005.

Goldstein, I. "Pleasure and Pain: Unconditional, Intrinsic Values." *Philosophy and Phenomenological Research* 50 (1989), 255–276.

Goodin, R. E. *On Settling*, Princeton, NJ: Princeton University Press, 2012.

Goodin, R. E. *Utilitarianism as a Public Philosophy*, Cambridge: Cambridge University Press, 1995.

Gosseries, A. and L. Meyer (eds.) *Intergenerational Justice*, Oxford: Oxford University Press, 2009.

Graham, P. "In Defense of Objectivism about Moral Obligation," *Ethics* 121 (2010), 88–115.

Green, T. H. and T. H. Grose, "Introduction" to D. Hume, *A Treatise of Human Nature*, vol. II, London: Longmans, Green, 1874.

Griffin, J. *Well-Being: Its Meaning, Measurement, and Moral Importance*, Oxford: Clarendon Press, 1986.

Grote, J. *An Examination of the Utilitarian Philosophy*, ed. J. B. Mayor, Cambridge: Deighton, Bell, 1870.

Gruzalski, B. "Foreseeable Consequence Utilitarianism," *Australasian Journal of Philosophy* 59 (1981), 163–176.

Guyer, P. *Kant on Freedom, Law, and Happiness*, Cambridge: Cambridge University Press, 2000.

Haakonssen, K. *Natural Law and Moral Philosophy: From Grotius to the Scottish Enlightenment*, Cambridge: Cambridge University Press, 1996.

Halévy, E. *The Growth of Philosophic Radicalism*, trans. M. Morris, Clifton, NJ: Augustus M. Kelley, 1972.

Hardin, G. "The Tragedy of the Commons," *Science* 162 (1968): 1243–1248.

Hardin, R. *Morality within the Limits of Reason*, Chicago, IL: University of Chicago Press, 1988.

Hare, R. M. "Could Kant Have been a Utilitarian?," *Utilitas* 5 (1993), 1–16.

Hare, R. M. *Moral Thinking: Its Levels, Method, and Point*, Oxford: Oxford University Press, 1981.

Hare, R. M. "Philosophy and Practice: Some Issues about War and Peace," in *Philosophy and Practice*, ed. A. P. Griffiths, Cambridge: Cambridge University Press, 1985, pp. 1–16 (Supplement to *Philosophy: Royal Institute of Philosophy Lecture Series* 18).

Hare, R. M. "Rules of War and Moral Reasoning," *Philosophy and Public Affairs* 1 (1972), 166–181.

Hare, R. M. *Sorting Out Ethics*, Oxford: Oxford University Press, 1997.

Hare, R. M. "The Structure of Ethics and Morals," in R. M. Hare, *Essays in Ethical Theory*, Oxford: Oxford University Press, 1989, pp. 175–190.

Harris, James, "The Epicurean in Hume," in *Epicurus in the Enlightenment*, eds. N. Leddy and A. S. Lifschitz, Oxford: Voltaire Foundation, 2009, pp. 105–118.

Harris, Jonathan, "Gay, John (1699–1745)," in *Oxford Dictionary of National Biography*, Oxford University Press, 2004; online edn., January 2008; at www.oxforddnb.com.ezproxy.lib.usf.edu/view/article/10474.

Harrison, J. "Utilitarianism, Universalization, and Our Duty to Be Just," *Proceedings of the Aristotelian Society* 53 (1953): 105–134.

Harrod, R. F. "Utilitarianism Revised," *Mind* 45 (1936), 137–156.

Harsanyi, J. C. "A Preference-Based Theory of Well-Being and a Rule-Utilitarian

Theory of Morality," in *Game Theory, Experience, Rationality: Foundations of Social Sciences, Economics and Ethics*, eds. W. Leinfellner and E. Köhler, Dordrecht: Kluwer, 1998, pp. 285–300.

Harsanyi, J. C. "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking," *Journal of Political Economy* 61 (1953), 434–435.

Harsanyi, J. C. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy* 63 (1955), 309–321.

Harsanyi, J. C. "Morality and the Theory of Rational Behaviour," in *Utilitarianism and Beyond*, eds. A. Sen and B. Williams, Cambridge: Cambridge University Press, 1982, pp. 39–62.

Harsanyi, J. C. "Rule Utilitarianism and Decision Theory," *Erkenntnis* 11 (1977), 25–53.

Harsanyi, J. C. "Some Epistemological Advantages of a Rule Utilitarian Position in Ethics," *Midwest Studies in Philosophy*, vol. VII, eds. P. A. French, T. E. Uehling, Jr., and H. K. Wettstein, Minneapolis, MN: University of Minnesota Press, 1982, pp. 389–402.

Haybron, D. *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*, Oxford: Oxford University Press, 2008.

Hayek, F. A. "The Use of Knowledge in Society," *American Economic Review* 35 (1945), 519–530.

Hazlitt, W. *The Complete Works of William Hazlitt*, 21 vols., ed. P. P. Howe, London: J. M. Dent, 1930–34.

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

Heathwood, C. "The Problem of Defective Desires," *Australasian Journal of Philosophy* 83 (2005), 487–504.

Heathwood, C. "The Reduction of Sensory Pleasure to Desire," *Philosophical Studies* 133 (2007), 23–44.

Heathwood, C. "Welfare," in *The Routledge Companion to Ethics*, ed. J. Skorupski, New York: Routledge, 2010, pp. 645–655.

Helvétius, C. *A Treatise on Man, His Intellectual Faculties and His Education [De L'Homme]*, London: B. Law and G. Robinson, 1777.

Helvétius, C. *De l'esprit, or, Essays on the Mind: And Its Several Faculties*, London: Dodsley, 1759.

Heyd, D. *Genethics: Moral Issues in the Creation of People*, Berkeley, CA: University of California Press, 1992.

Hildreth, R. *Theory of Morals: An Inquiry Concerning the Law of Moral Distinctions and the Variations and Contradictions of Ethical Codes*, Boston: C. C. Little & J. Brown, 1844.

Hill, Jr., T. E. "Assessing Moral Rules: Utilitarian and Kantian Perspectives," *Philosophical Issues* 15 (2005): 161–178.

Hill, Jr., T. E. "Kant on Imperfect Duty and Supererogation," in T. E. Hill, Jr., *Dignity and Practical Reason in Kant's Moral Theory*, Ithaca, NY: Cornell University Press, 1992, pp. 147–175.

Hobbes, T. *Leviathan*, ed. R. Tuck, Cambridge: Cambridge University Press, 1991.

Hodgson, D. H. *Consequences of Utilitarianism: A Study in Normative Ethics and Legal Theory*, Oxford: Clarendon Press, 1967.

Hoffman, D. *A Course of Legal Study; Respectfully Addressed to the Students of Law in the United States*, Baltimore, MD: Coale and Maxwell, 1817; 2nd edn., Baltimore, MD: Joseph Neal, 1836.

Hoffman, D. *Legal Outlines: Being the Substance of a Course of Lectures Now Delivering in the University of Maryland*, Baltimore, MD: E. J. Coale, 1829.

Holtug, N. "Prioritarianism," in *Egalitarianism: New Essays on the Nature and Value of Equality*, eds. N. Holtug and K. Lippert-Rasmussen, Oxford: Oxford University Press, 2006, pp. 125–156.

Hooker, B. "Does Moral Virtue Constitute a Benefit to the Agent?," in *How Should One Live? Essays on the Virtues* ed. R. Crisp, New York: Oxford University Press, 2003, pp. 141–155.

Hooker, B. "Fairness, Needs, and Desert," in *The Legacy of H. L. A. Hart: Legal, Political and Moral Philosophy*, eds. M. H. Kramer, C. Grant, B. Colburn, and A. Hatzistavrou, Oxford: Oxford University Press, 2008, pp. 181–199.

Hooker, B. *Ideal Code, Real World: A Rule-consequentialist Theory of Morality*, Oxford: Oxford University Press, 2000.

Hooker, B. "Reflective Equilibrium and Rule Consequentialism," in *Morality, Rules, and Consequences: A Critical Reader*, eds. B. Hooker, E. Mason, and D. E. Miller, Edinburgh: Edinburgh University Press, 2000, pp. 222–238.

Hooker, B. "Reply to Arneson and McIntyre," *Philosophical Issues* 15 (2005): 266–281.

Hooker, B. "Rule Consequentialism," in *The Stanford Encyclopedia of Philosophy* (spring 2011 edn.), ed. E. N. Zalta, at http://plato.stanford.edu/archives/spr2011/entries/consequentialism-rule.

Hooker, B. "When Is Impartiality Morally Appropriate?" in *Partiality and Impartiality: Morality, Special Relationships, and the Wider World*, eds. B. Feltham and J. Cottingham, Oxford: Oxford University Press, 2010, pp. 26–41.

Hooker, B. and G. Fletcher. "Variable versus Fixed-Rate Rule-Utilitarianism," *Philosophical Quarterly* 58 (2008), 344–352.

Howard-Snyder, F. "It's the Thought That Counts," *Utilitas* 17 (1997), 265–281.

Howard-Snyder, F. "The Rejection of Objective Consequentialism," *Utilitas* 9 (1997), 241–248.

Howard-Snyder, F. "Rule Consequentialism Is a Rubber Duck," *American Philosophical Quarterly* 30 (1993), 271–278.

Hudson, J. L. "Subjectivization in Ethics," *American Philosophical Quarterly* 26 (1989), 221–229.

Hume, D. *An Enquiry concerning the Principles of Morals*, ed. T. Beauchamp, Oxford: Oxford University Press, 2006.

Hume, L. J. *Bentham and Bureaucracy*, Cambridge: Cambridge University Press, 1981.

Hurka, T. *Perfectionism*, New York: Oxford University Press, 1993.

Hurka, T. "Value and Population Size," *Ethics* 93 (1983), 496–507.

Hursthouse, R. "Applying Virtue Ethics," in *Virtues and Reasons: Philippa Foot and Moral Theory*, ed. R. Hursthouse, G. Lawrence, and W. Quinn, Oxford: Oxford University Press, 1995, pp. 57–75.

Hursthouse, R. *On Virtue Ethics*, Oxford: Oxford University Press, 1999.

Hursthouse, R. "Practical Wisdom: A Mundane Account," *Proceedings of the Aristotelian Society* 106 (2006), 285–309.

Hursthouse, R. "The Virtuous Agent's Reasons: A Reply to Bernard Williams," in *Aristotle and Moral Realism*, ed. R. Heinaman, London: University College London Press, 1995, pp. 24–33.

Hutcheson, F. *An Inquiry into the Original of Our Ideas of Beauty and Virtue: in Two Treatises*, ed. W. Leidhold, Indianapolis, IN: Liberty Fund, 2004.

Ignatieff, M. *A Just Measure of Pain: The Penitentiary in the Industrial Revolution*

1750–1850, New York: Pantheon, 1978.

Irwin, T., *The Development of Ethics: A Historical and Critical Study, vol. III: From Kant to Rawls*, Oxford: Oxford University Press, 2009. See especially chapters 81–83, on Sidgwick.

Jackson, F. "A Probabilistic Approach to Moral Responsibility," *Studies in Logic and the Foundations of Mathematics* 114 (1986), 351–365.

Jackson, F. "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," *Ethics* 101 (1991), 461–482.

Jackson, F. "How Decision Theory Illuminates Assignments of Moral Responsibility," in *Intention in Law and Philosophy*, ed. N. Naffine, R. J. Owens, and J. Williams, Aldershot: Ashgate, 2001, pp. 19–36.

Jackson, F. and M. Smith, "Absolutist Moral Theories and Uncertainty," *Journal of Philosophy* 103 (2006), 267–283.

Jenyns, S. *A Free Inquiry into the Nature and Origin of Evil*, London: R. and J. Dodsley, 1757.

Jenyns, S. "Jeremy Bentham," *United States Magazine and Democratic Review* 8 (1840), 251–271.

Johnson, C. D. *Moral Legislation: A Legal-Political Model for Indirect Consequentialist Reasoning*, Cambridge: Cambridge University Press, 1991.

Johnson, R. "Kant's Moral Philosophy," in *The Stanford Encyclopedia of Philosophy* (summer 2012 edn.), ed. E. N. Zalta, at http://plato.stanford.edu/archives/sum2012/entries/kant-moral.

Kagan, S. "Evaluative Focal Points," in *Morality, Rules, and Consequences: A Critical Reader*, eds. B. Hooker, E. Mason, and D. E. Miller, Edinburgh: Edinburgh University Press, 2000, pp. 134–155.

Kagan, S. *The Geometry of Desert*, New York: Oxford University Press, 2012.

Kagan, S. *Normative Ethics*, Boulder, CO: Westview Press, 1998.

Kagan, S. "Well-Being as Enjoying the Good," *Philosophical Perspectives* 23 (2009), 253–272.

Kahn, L. "Rule Consequentialism and Scope," *Ethical Theory and Moral Practice* 15 (2012), 631–646.

Kahneman, D. *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux, 2011.

Kahneman, D. and A. Tversky, "Subjective Probability: A Judgment of Representativeness," *Cognitive Psychology* 3 (1972), 430–454.

Kant, I. *Critique of Practical Reason*, ed. M. Gregor, Cambridge: Cambridge University Press, 1997.

Kant, I. *Critique of Pure Reason*, eds. P. Guyer and A. W. Wood, Cambridge: Cambridge University Press, 1998.

Kant, I. *Groundwork of the Metaphysics of Morals*, eds. M. Gregor and J. Timmermann, Cambridge: Cambridge University Press, 2012.

Kant, I. *Lectures on Ethics*, eds. P. L. Heath and J. B. Schneewind, Cambridge: Cambridge University Press, 1997.

Kant, I. *The Metaphysics of Morals*, ed. M. Gregor, Cambridge: Cambridge University Press, 1996.

Kapur, N. B. "Why It Is Wrong to Be Always Guided by the Best: Consequentialism and Friendship," *Ethics* 101 (1991), 483–504.

Kawall, J. "The Experience Machine and Mental State Theories of Well-Being," *Journal of Value Inquiry* 33 (1999), 381–387.

Keller, S. "Welfare and the Achievement of Goals," *Philosophical Studies* 121 (2004), 27–41.

Kelly, P. "Utilitarianism and Distributive Justice: The Civil Law and the Foundations of Bentham's Economic Thought," *Utilitas* 1 (1989), 62–81.

Kerner, G. C. "The Immortality of Utilitarianism and the Escapism of Rule-Utilitarianism," *Philosophical Quarterly* 21 (1971), 36–50.

King, P. J. *Utilitarian Jurisprudence in America: The Influence of Bentham and Austin on American Legal Thought in the Nineteenth Century*, New York and London: Garland, 1986.

Korsgaard, C. M. "Kant's Formula of Universal Law," in C. M. Korsgaard, *Creating the Kingdom of Ends*, Cambridge: Cambridge University Press, 1996, pp. 77–105.

Korsgaard, C. M. *The Sources of Normativity*, Cambridge: Cambridge University Press, 1996.

Kraut, R. "Desire and the Human Good," *Proceedings and Addresses of the American Philosophical Association* 68 (1994), 39–54.

Kumar, R. "Who Can Be Wronged?" *Philosophy and Public Affairs* 31 (2003), 99–118.

Kymlicka, W. *Contemporary Political Philosophy: An Introduction*, 2nd edn., Oxford: Oxford University Press, 2002.

Lackey, D. *The Ethics of War and Peace*, Englewood Cliffs, NJ: Prentice-Hall, 1989.

Law, E. "Morality and Religion," in W. King, *An Essay on the Origin of Evil*, 4th edn., ed. E. Law, Cambridge: 1758, pp. xliii–lii.

Law, E. "The Nature and Obligations of Man, as a Sensible and Rational Being," in W. King, *An Essay on the Origin of Evil*, 4th edn., ed. E. Law, Cambridge: 1758, pp. liii–lx.

Layard, R. *Happiness: Lessons from a New Science*, New York: Penguin, 2005.

Lazari-Radek, K. de and P. Singer, "Secrecy in Consequentialism: A Defence of Esoteric Morality," *Ratio* 23 (2010), 34–58.

Legaré, H. S. "Jeremy Bentham and the Utilitarians," *Southern Review* 7 (1831), 261–296.

Lemos, N. "Indeterminate Value, Basic Value, and Summation," in *The Good, The Right, Life and Death: Essays in Honor of Fred Feldman*, eds. K. McDaniel, J. Raibley, R. Feldman, and M. Zimmerman, Burlington, VT: Ashgate Publishing Co., 2006, pp. 71–82.

Lenman, J. "Consequentialism and Cluelessness," *Philosophy and Public Affairs* 29 (2000), 342–370.

Lewis, C. I. *The Ground and Nature of the Right*, New York: Columbia University Press, 1955.

Lewis, D. *On the Plurality of Worlds*, Malden, MA: Blackwell Publishers, 1986.

Lieberman, D. "Bentham on Codification," in J. Bentham, *Selected Writings*, edited with an introduction by S. G. Engelmann, with essays by P. Schofield, D. Lieberman, J. Pitts, and M. Canuel, New Haven, CT and London: Yale University Press, 2011, pp. 460–477.

Lively, J. and J. Rees (eds.) *Utilitarian Logic and Politics: James Mill's "Essay on Government," Macaulay's Critique and the Ensuing Debate*, Oxford: Clarendon Press, 1978.

Livingston, E. *A System of Penal Law for the State of Louisiana*, Philadelphia, PA: James Kay, Jun. & Co., 1833.

Locke, J. *An Essay concerning Human Understanding*, ed. P. H. Nidditch, Oxford: Oxford University Press, 1975.

Locke, J. *Political Essays*, ed. M. Goldie, Cambridge: Cambridge University Press, 1997.

Locke, J. *Two Treatises of Government*, ed. P. Laslett, Cambridge: Cambridge University Press, 1988.

Lockhart, T. *Moral Uncertainty and Its Consequences*, New York: Oxford University Press, 2000.

Louden, R. B. "On Some Vices of Virtue Ethics," *American Philosophical Quarterly* 21 (1984), 227–236.

Lukas, M. "Desire Satisfactionism and the Problem of Irrelevant Desires," *Journal of Ethics & Social Philosophy* 4 (2012), 1–24.

Luño, A.-E. P. "Jeremy Bentham and Legal Education in the University of Salamanca during the Nineteenth Century," *Bentham Newsletter* 5 (1981), 44–54.

Lyons, D. *Forms and Limits of Utilitarianism*, Oxford: Clarendon Press, 1965.

Lyons, D. "Mill's Theory of Morality," in D. Lyons, *Rights, Welfare, and Mill's Moral Theory*, Oxford: Oxford University Press, 1994, pp. 47–65.

Lyons, D. "Utility and Rights," in D. Lyons, *Rights, Welfare, and Mill's Moral Theory*, Oxford: Oxford University Press, 1994, pp. 147–175.

Mackie, J. L. "Cooperation, Competition and Moral Philosophy," in *Cooperation and Competition in Humans and Animals*, ed. A. M. Colman, Wokingham: van Nostrand Reinhold, 1982, pp. 271–284.

Mackie, J. L. "The Law of the Jungle: Moral Alternatives and Principles of Evolution," *Philosophy* 53 (1978), 455–464.

Mackintosh, J. *Dissertation on the Progress of Ethical Philosophy, Chiefly during the Seventeenth and Eighteenth Centuries*, 3rd edn., Edinburgh: A. & C. Black, 1862.

McCloskey, H. J. "A Non-Utilitarian Approach to Punishment," *Inquiry* 8 (1965), 249–263.

McCloskey, H. J. "Utilitarianism: Two Difficulties," *Philosophical Studies* 24 (1973), 62–63.

McKennon, T. L. "Benthamism in Santander's Colombia," *Bentham Newsletter* 5 (1981), 29–43.

McKerlie, D. "Equality and Priority," *Utilitas* 6 (1994), 25–42.

Mādhava Āchārya *The Sarva-Darśana-Samgraha*, trans. E. B. Cowell and A. E.

Gough, London: Trübner, 1882.

Mandelbaum, M. "On Interpreting Mill's *Utilitarianism*," *Journal of the History of Philosophy* 6 (1968), pp. 35–46.

Marx, K. Capital: A Critique of Political Economy, vol. I, in K. Marx and F. Engels, *Collected Works* vol. xxxv, New York: International Publishers, 1975.

Mason, E. "Consequentialism and the 'Ought Implies Can' Principle," *American Philosophical Quarterly* 40 (2003), 319–331.

Mason, E. "Objectivism and Prospectivism about Rightness," *Journal of Ethics & Social Philosophy* 7 (2013); at www.jesp.org/PDF/objectivism_and_propectivism_about_rightness.pdf.

Mason, E. *Subjective Consequentialism*, manuscript.

Mele, A. "Agents' Abilities," *Noûs* 37 (2003), 447–470.

Mendola, J. *Goodness and Justice: A Consequentialist Moral Theory*, New York: Cambridge University Press, 2006.

Mill, J. *A Fragment on Mackintosh*, London: Longmans, 1835.

Mill, J. *Analysis of the Phenomena of the Human Mind* (2 vols.), 2nd edn., ed. J. S. Mill, New York: Augustus M. Kelley, 1967.

Mill, J. *Commerce Defended: An Answer to the Arguments by which Mr. Spence, Mr. Corbett, and Others, Have Attempted to Prove that Commerce Is Not a Source of National Wealth*, London: C. and R. Baldwin, 1808.

Mill, J. *"Law of Nations," in Supplement to the Encyclopaedia Britannica*, London: J. Innis, 1825.

Mill, J. *Political Writings*, ed. T. Ball, Cambridge: Cambridge University Press, 1992.

Mill, J. S. *The Collected Works of John Stuart Mill*, ed. J. M. Robson, 33 vols., Toronto: University of Toronto Press, 1963–91.

Miller, D. *Principles of Social Justice*, Cambridge, MA: Harvard University Press, 1999.

Miller, D. E. "Actual-Consequence Act Utilitarianism and the Best Possible Humans," *Ratio* 16 (2003), 49–62.

Miller, D. E. "Hooker's Use and Abuse of Reflective Equilibrium," in *Morality, Rules, and Consequences: A Critical Reader*, eds. B. Hooker, E. Mason, and D. E. Miller, Edinburgh: Edinburgh University Press, 2000, pp. 156–178.

Miller, D. E. *J. S. Mill: Moral, Social and Political Thought*, Cambridge: Polity, 2010.

Miller, D. E. "Mill, Rule Utilitarianism, and the Incoherence Objection," in *John Stuart Mill and the Art of Life*, eds. B. Eggleston, D. E. Miller, and D. Weinstein, New York: Oxford University Press, 2011, pp. 94–116.

Miller, R. B. "Actual Rule Utilitarianism," *Journal of Philosophy* 106 (2009), 5–28.

Montmarquet, J. "Zimmerman on Culpable Ignorance," *Ethics* 109 (1999), 842–845.

Moore, G. E. *Ethics*, ed. W. H. Shaw, Oxford: Oxford University Press, 2005.

Moore, G. E. *Principia Ethica*, Cambridge: Cambridge University Press, 1903.

Moore, J. "Hume and Hutcheson," in *Hume and Hume's Connexions*, eds. M. A. Stewart and J. P. Wright, Edinburgh: Edinburgh University Press, 1994, pp. 23–57.

Moore, J. "The Eclectic Stoic, the Mitigated Sceptic," in *New Essays on David Hume*, eds. E. Mazza and E. Ronchetti, Milan: FrancoAngeli, 2007, pp. 133–169.

Mulgan, T. *The Demands of Consequentialism*, Oxford: Oxford University Press, 2001.

Mulgan, T. *Ethics for a Broken World: Imagining Philosophy after Catastrophe*, Acumen, Durham, UK, 2011.

Mulgan, T. *Future People: A Moderate Consequentialist Account of Our Obligations to Future Generations*, Oxford: Oxford University Press, 2006.

Mulgan, T. "One False Virtue of Rule Consequentialism, and One New Vice" *Pacific Philosophical Quarterly* 77 (1996), 362–373.

Mulgan, T. "Utilitarianism for a Broken World," presented at ISUS XII (12th conference of the International Society for Utilitarian Studies), New York University, New York, NY, August 11, 2012.

Mulgan, T. "What Is Good for the Distant Future? The Challenge of Climate Change for Utilitarianism," forthcoming in *God, The Good, and Utilitarianism: Perspectives on Peter Singer*, ed. J. Perry, Cambridge: Cambridge University Press, 2014.

Murphy, L. B. *Moral Demands in Nonideal Theory*, New York: Oxford University Press, 2000.

Myers, F. W. H. *Fragments of Prose and Poetry*, ed. E. Myers, London: Longmans, Green, 1904.

Nagel, T. "Death," *Noûs* 4 (1970), 73–80.

Nagel, T. *Equality and Partiality*, New York: Oxford University Press, 1991.

Nathanson, S. *Terrorism and the Ethics of War*, Cambridge: Cambridge University Press, 2010.

Neal, J. *Principles of Legislation: . . . with notes and a biographical notice of Jeremy Bentham and of M. Dumont*, Boston, MA: Wells and Lilly, 1830.

Neal, J. *Wandering Recollections of a Somewhat Busy Life*, Boston, MA: Roberts Brothers, 1869.

Ng, Y.-K. "What Should We Do about Future Generations? Impossibility of Parfit's Theory X," *Economics and Philosophy* 5 (1989), 235–253.

Nietzsche, F. *Beyond Good and Evil: Prelude to a Philosophy of the Future*, ed. R.-P. Horstmann and J. Norman, Cambridge: Cambridge University Press, 2002.

Nietzsche, F. Twilight of the Idols, or How to Philosophize with a Hammer, in F. Nietzsche, *The Anti-Christ, Ecco Homo, Twilight of the Idols: And Other Writings*, eds. A. Ridley and J. Norman, Cambridge: Cambridge University Press, 2005, pp. 153–229.

Norcross, A. "The Scalar Approach to Utilitarianism," *The Blackwell Guide to Mill's Utilitarianism*, ed. H. R. West, Malden, MA, and Oxford: Blackwell Publishing, 2006, pp. 217–232.

Nordhaus, W. *The Challenge of Global Warming: Economic Models and Environmental Policy*, available at nordhaus.econ.yale.edu/dice_mss_072407_all.pdf (2007).

Norton, D. F. *David Hume: Common-Sense Moralist, Sceptical Metaphysician*, Princeton, NJ: Princeton University Press, 1982.

Nowell-Smith, P. H. *Ethics*, Harmondsworth: Penguin, 1954.

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

Nussbaum, M. "Adaptive Preferences and Women's Options," *Economics and Philosophy* 17 (2001), 67–88.

Oddie, G. and P. Menzies, "An Objectivist's Guide to Subjective Value," *Ethics* 102 (1992), 512–533.

Olsaretti, S. "Distributive Justice and Compensatory Desert," in *Desert and Justice*, ed. S. Olsaretti, Oxford: Oxford University Press, 2003, pp. 187–204.

O'Sullivan, J. L. *Report in Favor of the Abolition of the Punishment of Death by Law*, 2nd edn., New York: J. & H. G. Langley, 1841.

Overvold, M. C. "Self-Interest and Getting What You Want," in *The Limits of*

Utilitarianism, eds. H. B. Miller and W. H. Williams, Minneapolis, MN: University of Minnesota Press, 1982, pp. 186–194.

Packe, M. S. *The Life of John Stuart Mill*, London: Secker & Warburg, 1954.

Paley, W. *The Principles of Moral and Political Philosophy*, Indianapolis, IN: Liberty Fund, 2002.

Palmer, D. E. "On the Viability of a Rule Utilitarianism," *Journal of Value Inquiry* 33 (1999), 31–42.

Parfit, D. "Equality and Priority," *Ratio* 10 (1997), 202–221.

Parfit, D. "Equality or Priority?" in *The Ideal of Equality*, eds. M. Clayton and A. Williams, Basingstoke, UK: Palgrave Macmillan, 2000, pp. 81–125.

Parfit, D. *On What Matters*, 2 vols., Oxford: Oxford University Press, 2011.

Parfit, D. "Overpopulation and the Quality of Life," in *Applied Ethics*, ed. P. Singer, Oxford: Oxford University Press, 1986, pp. 145–164.

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

Peonidis, F. "Bentham and the Greek Revolution: New Evidence," *Journal of Bentham Studies* 11 (2009), available at http://ojs.lib.ucl.ac.uk/index.php/jbs/article/view/60.

Persson, I. "A Consequentialist Distinction between What We Ought to Do and Ought to Try," *Utilitas* 20 (2008), 348–355.

Persson, I. "Universalizability and the Summing of Desires," *Theoria* 55 (1989), 159–170.

Petersson, B. "The Second Mistake in Moral Mathematics is not about the Worth of Mere Participation," *Utilitas* 16 (2004), 288–315.

Pettit, P. "Satisficing Consequentialism," part II, *Proceedings of the Aristotelian Society* supplementary volume 58 (1984), 165–176.

Pettit, P. and M. Smith, "Global Consequentialism," in *Morality, Rules, and Consequences: A Critical Reader*, eds. B. Hooker, E. Mason, and D. E. Miller, Edinburgh: Edinburgh University Press, 2000, pp. 121–133.

Phillips, D. *Sidgwickian Ethics*, Oxford: Oxford University Press, 2011.

"*Plan of Parliamentary Reform* . . . by Jeremy Bentham," *Edinburgh Review, or Critical Journal* 31 (1818), 165–203.

"*Plan of Parliamentary Reform* . . . by Jeremy Bentham," *Quarterly Review* 18 (1817), 128–135.

Plato, Euthyphro, in Plato, *Complete Works*, ed. J. Cooper, Indianapolis, IN: Hackett, 1997, pp. 1–16.

Pojman, L. P. and O. MacLeod, *What Do We Deserve? A Reader on Justice and Desert*, New York: Oxford University Press, 1999.

Portmore, D. *Commonsense Consequentialism: Wherein Morality Meets Rationality*, Oxford: Oxford University Press, 2011.

Posner, R. *The Problems of Jurisprudence*, Cambridge, MA: Harvard University Press, 1990.

Prichard, H. A. "Duty and Ignorance of Fact," in his *Moral Writings*, ed. J. MacAdam, Oxford: Clarendon Press, 2002, pp. 84–101.

Quinn, M. "A Failure to Reconcile the Irreconcilable? Security, Subsistence and Equality in Bentham's Writings on the Civil Code and on the Poor Laws," *History of Political Thought* 29 (2008), 320–343.

Rabinowicz, W. "Preference Utilitarianism by Way of Preference Change?" in *Preference Change: Approaches from Philosophy, Economics and Psychology*, eds. T. Grüne-Yanoff and S. O. Hansson, Dordrecht: Springer, 2009, pp. 185–206.

Rabinowicz, W. and B. Strömberg, "What If I Were in His Shoes? On Hare's Argument for Preference Utilitarianism," *Theoria* 62 (1996), 95–123.

Rachels, S. "A Set of Solutions to Parfit's Problems," *Noûs* 35 (2001), 214–238.

Radzinowicz, L. *A History of English Criminal Law and its Administration from 1750*, 3 vols., London: Stevens, 1948.

Raibley, J. "Well-Being and the Priority of Values," *Social Theory and Practice* 36 (2010), 593–620.

Railton, P. "Alienation, Consequentialism, and the Demands of Morality," *Philosophy and Public Affairs* 13 (1984), 134–171.

Railton, P. "Facts and Values," *Philosophical Topics* 14 (1986), 5–31.

Railton, P. "Naturalism and Prescriptivity," *Social Philosophy and Policy* 7 (1989), 151–174.

Rawls, J. *A Theory of Justice*, Cambridge, MA: Belknap Press, 1971.

Rawls, J. "Justice as Fairness," *Philosophical Review* 67 (1958), 164–194.

Rawls, J. "Two Concepts of Rules," in J. Rawls, *Collected Papers*, ed. S. Freeman, Cambridge, MA: Harvard University Press, 1999, pp. 20–46.

Raz, J. *The Morality of Freedom*, Oxford: Clarendon Press, 1986.

Regan, D. *Utilitarianism and Co-operation*, Oxford: Clarendon Press, 1980.

Rhoads, S. E. *The Economist's View of the World: Government, Markets, and Public Policy*, Cambridge: Cambridge University Press, 1985.

Ridge, M. "Introducing Variable-Rate Rule-Utilitarianism," *Philosophical Quarterly* 56 (2006), 242–253.

Riley, J. "Defending Rule Utilitarianism," in *Morality, Rules, and Consequences: A Critical Reader*, eds. B. Hooker, E. Mason, and D. E. Miller, Edinburgh: Edinburgh University Press, 2000, pp. 40–70.

Rivers, I. *Reason, Grace, and Sentiment: A Study of the Language of Religion and Ethics in England 1660–1780, vol. 2: Shaftesbury to Hume*, Cambridge: Cambridge University Press, 2000.

Roberts, M. "A New Way of Doing the Best That We Can: Person-Based Consequentialism and the Equality Problem," *Ethics* 112 (2001), 315–350.

Roberts, M. and D. Wasserman (eds.) *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem*, Dordrecht: Springer, 2009.

Robson, J. M. "Textual Introduction," in J. S. Mill, *Collected Works*, vol. x, pp. cxv–cxxxix.

Rosati, C. S. "Persons, Perspectives, and Full Information Accounts of the Good," *Ethics* 105 (1995), 296–325.

Rosen, F. *Classical Utilitarianism from Hume to Mill*, London: Routledge, 2003.

Ross, I. C. "Was Berkeley a Jacobite? Passive Obedience Revisited," *Eighteenth-Century Ireland* 20 (2005), 17–30.

Ross, W. D. *Foundations of Ethics*, Oxford: Clarendon Press, 1939.

Ross, W. D. *The Right and the Good*, Oxford: Clarendon Press, 1930.

Russell, D. C. *Practical Intelligence and the Virtues*, Oxford: Oxford University Press, 2009.

Rutherforth, T. *An Essay on the Nature and Obligations of Virtue*, Cambridge: J. Bentham, 1744.

Rutherforth, T. *Institutes of Natural Law: Being the Substance of a Course of Lectures on Grotius De Jure Belli Et Pacis, Two Volumes*, Cambridge: J. Bentham, 1754–1756.

Ryberg, J. and T. Tännsjö (eds.) *The Repugnant Conclusion: Essays on Population*

Ethics, Dordrecht: Kluwer, 2004.

Sandel, M. J. "How Markets Crowd out Morals," *Boston Review*, May/June 2012; at www.bostonreview.net/BR37.3/ndf_michael_j_sandel_markets_morals.php.

Sandel, M. J. *Justice: What's the Right Thing to Do?* New York: Farrar, Straus and Giroux, 2009.

Sandel, M. J. "What Isn't for Sale?" *The Atlantic*, April 2012; at www.theatlantic.com/magazine/archive/2012/04/what-isnt-for-sale/308902/.

Scanlon, T. M. "Contractualism and Utilitarianism," in *Utilitarianism and Beyond*, eds. A. Sen and B. Williams, Cambridge: Cambridge University Press, 1982, pp. 103–128.

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

Scarre, G. *Utilitarianism*, London: Routledge, 1996.

Scheffler, S. *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*, Oxford: Oxford University Press, 1982.

Schmidtz, D. "A Place for Cost–Benefit Analysis," *Philosophical Issues* 11 (2001), 148–171.

Schneewind, J. B. "Introduction," in J. S. Mill, *Mill's Ethical Writings*, ed. J. B. Schneewind, London: Collier-Macmillan and New York: Collier, 1965.

Schneewind, J. B. *The Invention of Autonomy*, Cambridge: Cambridge University Press, 1998.

Schneewind, J. B. "The Misfortunes of Virtue," *Ethics* 101 (1990), 42–63.

Schneewind, J. B. *Sidgwick's Ethics and Victorian Moral Philosophy*, Oxford: Clarendon Press, 1977.

Schofield, P. *Utility and Democracy: The Political Thought of Jeremy Bentham*, Oxford: Oxford University Press, 2006.

Schueler, G. F. "Some Reasoning about Preferences," *Ethics* 95 (1984), 78–80.

Schultz, B. (ed.) *Essays on Henry Sidgwick*, Cambridge: Cambridge University Press, 1992.

Schultz, B., *Henry Sidgwick – Eye of the Universe: An Intellectual Biography*, Cambridge: Cambridge University Press, 2004.

Sedgwick, A. *A Discourse on the Studies of the University*, Leicester: Leicester

University Press, 1969.

Semple, J. *Bentham's Prison: A Study of the Panopticon Penitentiary*, Oxford: Clarendon Press, 1993.

Sen, A. "Utilitarianism and Welfarism," *Journal of Philosophy* 76 (1979), 463–489.

Sen, A. "Utility: Ideas and Terminology," *Economics and Philosophy* 7 (1991), 277–283.

Sepielli, A. "What to Do When You Don't Know What to Do," *Oxford Studies in Metaethics*, vol. IV, ed. Russ Shafer-Landau, Oxford: Oxford University Press, 2009, pp. 5–28.

Shaftesbury, Earl of (A. A. Cooper), *Characteristicks of Men, Manners, Opinions, Times*, Indianapolis, IN: Liberty Fund, 2001.

Shaw, W. H. "Consequentialism, War, and National Defense," forthcoming in *Journal of International Political Theory* 10 (2014).

Shaw, W. H. *Contemporary Ethics: Taking Account of Utilitarianism*, Oxford: Blackwell, 1999.

Shaw, W. H. "Just War Theory," in *Social and Personal Ethics*, ed. W. H. Shaw, 8th edn., Belmont, CA: Wadsworth/Cengage, 2014, pp. 341–347.

Shaw, W. H. "Utilitarianism and Recourse to War," *Utilitas* 23 (2011), 380–401.

Shrader-Frechette, K. "Parfit and Mistakes in Moral Mathematics," *Ethics* 98 (1987), 50–60.

Sidgwick, H. *The Elements of Politics*, 3rd edn., London: Macmillan, 1908.

Sidgwick, H. *Essays on Ethics and Method*, ed. M. G. Singer, Oxford: Clarendon Press, 2000.

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

Sidgwick, H. *Miscellaneous Essays and Addresses*, London: Macmillan, 1904.

Sidgwick, H. "The Morality of Strife," in H. Sidgwick, *Practical Ethics: A Collection of Addresses and Essays*, New York: Oxford University Press, 1998, pp. 47–62.

Sidgwick, H. *Outlines of the History of Ethics*, London: Macmillan, 1886.

Singer, P. *Animal Liberation: The Definitive Classic of the Animal Movement*, New York: Harper Perennial, 2009.

Singer, P. "Famine, Affluence, and Morality," *Philosophy and Public Affairs* 1 (1972),

229–243.

Singer, P. *Marx*, New York: Farrar, Straus and Giroux, 1980.

Singer, P. *Practical Ethics*, 3rd edn., Cambridge: Cambridge University Press, 2011.

Sinnott-Armstrong, W. "Consequentialism," in *The Stanford Encyclopedia of Philosophy* (winter 2012 edn.), ed. E. N. Zalta, at http://plato.stanford.edu/archives/win2012/entries/consequentialism.

Skorupski, J. (ed.) *The Cambridge Companion to Mill*, Cambridge: Cambridge University Press, 1998.

Skorupski, J. *John Stuart Mill*, London: Routledge & Kegan Paul, 1989.

Skorupski, J., "The Place of Utilitarianism in Mill's Philosophy," in *The Blackwell Guide to Mill's Utilitarianism*, ed. H. R. West, Malden, MA, and Oxford: Blackwell Publishing, 2006, pp. 45–59.

Slote, M. A. *Beyond Optimizing: A Study of Rational Choice*, Cambridge, MA: Harvard University Press, 1989.

Slote, M. A. *Common-Sense Morality and Consequentialism*, London: Routledge & Kegan, 1985.

Slote, M. A. *From Morality to Virtue*, New York: Oxford University Press, 1992.

Slote, M. A. *Morals from Motives*, Oxford: Oxford University Press, 2003.

Slote, M. A. "Satisficing Consequentialism," part I, *Proceedings of the Aristotelian Society* supplementary volume 58 (1984), 139–163.

Slote, M. A. "Two Views of Satisficing," in *Satisficing and Maximizing: Moral Theorists on Practical Reason*, ed. M. Byron, Cambridge: Cambridge University Press, 2004, pp. 14–29.

Smart, J. J. C. *An Outline of a System of Utilitarian Ethics*, Melbourne: Melbourne University Press, 1961.

Smart, J. J. C. "An Outline of a System of Utilitarian Ethics," in J. J. C. Smart and B. Williams, *Utilitarianism: For and Against*, Cambridge: Cambridge University Press, 1973, pp. 1–74.

Smart, J. J. C. "Extreme and Restricted Utilitarianism," *Philosophical Quarterly* 6 (1956), 344–354.

Smith, A. *An Inquiry into the Nature and Causes of the Wealth of Nations*, 2 vols., eds. R. H. Campbell and A. S. Skinner, Oxford: Oxford University Press, 1979.

Smith, A. *The Theory of Moral Sentiments*, eds. D. D. Raphael and A. L. Macfie, Oxford: Oxford University Press, 1979.

Smith, G. "Utilitarianism," in *Oxford Reader's Companion to Dickens*, ed. P. Schlicke, Oxford: Oxford University Press, 1999, pp. 581–582.

Smith, H. M. "Culpable Ignorance," *Philosophical Review* 92 (1983), 543–571.

Smith, H. M. "Deciding How to Decide: Is There a Regress Problem?" in *Essays in the Foundations of Decision Theory*, ed. M. Bacharach and S. Hurley, Oxford: Basil Blackwell, Inc., 1991, pp. 194–219.

Smith, H. M. "Making Moral Decisions," *Noûs* 22 (1988), 89–108.

Smith, H. M. "Measuring the Consequences of Rules," *Utilitas* 22 (2010), 413–433.

Smith, H. M. "The 'Prospective View' of Obligation," *Journal of Ethics & Social Philosophy* 5 (2011); at www.jesp.org/articles/download/ProspectiveView.pdf.

Smith, H. M. "Subjective Rightness," *Social Philosophy and Policy* 27:2 (2010), 64–110.

Smith, H. M. "Two-Tier Moral Codes," *Social Philosophy and Policy* 7:1 (1989), 112–132.

Smith, H. M. "Varieties of Moral Worth and Moral Credit," *Ethics* 101 (1991), 279–303.

Sobel, D. "Full Information Accounts of Well-Being," *Ethics* 104 (1994), 784–810.

Spencer, H. *The Data of Ethics*, London: Williams and Norgate, 1879.

Spencer, H. *The Principles of Psychology*, 2nd edn., New York: D. Appleton, 1871.

Spinoza, B. *Ethics*, in B. Spinoza, *The Collected Works of Spinoza*, ed. and trans. E. Curley, vol. I, pp. 408–617.

Stephen, J. F. *Horæ Sabbaticæ*, 3 vols., London: Macmillan, 1892.

Stephen, L. *History of English Thought in the Eighteenth Century*, vol. II, London: Smith, Elder, 1902.

Stephen, L. *The Science of Ethics*, London: Smith, Elder, 1882.

Stern, N. *Stern Review: The Economics of Climate Change;* at http://webarchive.nationalarchives.gov.uk/+/ http:/www.hm-treasury.gov.uk/sternreview_index.htm, 2006.

Stocker, M. "The Schizophrenia of Modern Ethical Theories," *Journal of Philosophy*

73 (1976), 453–466.

Stone, M. "Dickens, Bentham, and the Fictions of the Law: A Victorian Controversy and Its Consequences," *Victorian Studies* 29 (1985), 125–154.

Streumer, B. "Can Consequentialism Cover Everything?" *Utilitas* 15 (2003), 237–247.

Strong, T. B. *Friedrich Nietzsche and the Politics of Transformation*, expanded edn., Berkeley, CA: University of California Press, 1988.

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

Temkin, L. "Intransitivity and the Mere Addition Paradox," *Philosophy and Public Affairs* 16 (1987), 138–187.

Temkin, L. "Theory of Legislation, by Jeremy Bentham," *North American Review* 51 (1840), 384–396.

Thomas, W. *The Philosophic Radicals: Nine Studies in Theory and Practice, 1817–1841*, Oxford: Clarendon Press, 1979.

Thompson, W. *Appeal of One Half of the Human Race Women against the Pretensions of the other Half Men, to Retain Them in Political, and Thence in Civil and Domestic, Slavery; in Reply to a Paragraph of Mr. Mill's Celebrated "Article on Government,"* New York: Burt Franklin, 1970.

Tiberius, V. and A. Plakias, "Well-Being," in *The Moral Psychology Handbook*, eds. J. M. Doris and the Moral Psychology Research Group, Oxford: Oxford University Press, 2010, pp. 402–432.

Tierney, B. *The Idea of Natural Rights: Studies on Natural Rights, Natural Law, and Church Law, 1150–1625*, Atlanta, GA: Scholars Press, 1997.

Timmermann, J. "Good But Not Required? – Assessing the Demands of Kantian Ethics," *Journal of Moral Philosophy* 2 (2005), 9–27.

Timmermann, J. *Kant's Groundwork of the Metaphysics of Morals: A Commentary*, Cambridge: Cambridge University Press, 2007.

Timmermann, J. "When the Tail Wags the Dog: Animal Welfare and Indirect Duty in Kantian Ethics," *Kantian Review* 10 (2005), 128–149.

Timmermann, J. "Why Kant Could Not Have Been a Utilitarian," *Utilitas* 17 (2005), 243–264.

Timmons, M. *Moral Theory: An Introduction*, Lanham, MD: Rowman & Littlefield, 2002.

Tuck, R. *Natural Rights Theories: Their Origin and Development*, Cambridge: Cambridge University Press, 1982.

Tucker, A. *The Light of Nature Pursued*, 5 vols., London: 1768.

Tully, J. *A Discourse on Property: John Locke and His Adversaries*, Cambridge: Cambridge University Press, 1980.

Unger, P. *Living High and Letting Die*, Oxford: Oxford University Press, 1996.

Urmson, J. O. "The Interpretation of the Moral Philosophy of J. S. Mill," *Philosophical Quarterly* 3 (1953), 33–39.

Vallentyne, P. "Against Maximizing Act Consequentialism," in *Contemporary Debates in Moral Theory*, ed. J. Dreier, Malden, MA: Blackwell Publishing, 2006, pp. 21–37.

Vallentyne, P. "Utilitarianism and Infinite Utility," *Australasian Journal of Philosophy* 71 (1993), 212–217.

Vallentyne, P. and S. Kagan, "Infinite Value and Finitely Additive Value Theory," *Journal of Philosophy* 94 (1997), 5–26.

von Wright, G. H. *The Varieties of Goodness*, London: Routledge & Kegan Paul, 1963.

Waldron, J. *God, Locke, and Equality: Christian Foundations in Locke's Political Thought*, Cambridge: Cambridge University Press, 2002.

Wall, G. "Mill on Happiness as an End," *Philosophy* 57 (1982), 537–541.

Walzer, M. *Just and Unjust Wars: A Moral Argument with Historical Illustrations*, 3rd edn., New York: Basic Books, 2000.

Warke, T. "Multi-Dimensional Utility and the Index Number Problem: Jeremy Bentham, J. S. Mill, and Qualitative Hedonism," *Utilitas* 12 (2000), 176–203.

West, H. R. *An Introduction to Mill's Utilitarian Ethics*, Cambridge: Cambridge University Press, 2004.

West, H. R. *Mill's Utilitarianism: A Reader's Guide*, London: Continuum, 2007.

Weymark, J. A. "A Reconsideration of the Harsanyi–Sen Debate on Utilitarianism," in *Interpersonal Comparisons of Well-Being*, eds. J. Elster and J. E. Roemer, Cambridge: Cambridge University Press, 1991, pp. 255–320.

Whewell, W. *Lectures on the History of Moral Philosophy in England and Additional Lectures on History of Moral Philosophy*, Bristol, UK: Thoemmes Press, 1990.

Wiggins, D. "Claims of Need," in D. Wiggins, *Needs, Value, Truth: Essays in the*

Philosophy of Value, 2nd edn., Oxford: Blackwell, 1991, pp. 1–57.

Wiland, E. "Monkeys, Typewriters, and Objective Consequentialism," *Ratio* 18 (2005), 352–360.

Williams, B. "A Critique of Utilitarianism," in J. J. C. Smart and B. Williams, *Utilitarianism: For and Against*, Cambridge: Cambridge University Press, 1973, pp. 75–150.

Williams, B. *Ethics and the Limits of Philosophy*, Cambridge, MA: Harvard University Press, 1985.

Williams, G. "J. S. Mill and Political Violence," *Utilitas* 1 (1989), 102–111.

Wolf, S. *Meaning in Life and Why It Matters*, Princeton, NJ: Princeton University Press, 2010.

Wolf, S. "Moral Saints," *Journal of Philosophy* 79 (1982), 419–439.

Wootton, D. "Helvétius: From Radical Enlightenment to Revolution," *Political Theory* 28 (2000), 307–336.

Yasukawa, R. "James Mill on Peace and War," *Utilitas* 3 (1991), 179–197.

Zimmerman, M. J. "Is Moral Obligation Objective or Subjective?" *Utilitas* 18 (2006), 329–361.

Zimmerman, M. J. *Living with Uncertainty: The Moral Significance of Ignorance*, Cambridge: Cambridge University Press, 2008.

Zimmerman, M. J. "Moral Responsibility and Ignorance," *Ethics* 107 (1997), 410–426.

Zwolinski, M. "The Ethics of Price-Gouging," *Business Ethics Quarterly* 18 (2008), 347–378.

Zwolinski, M. and D. Schmidtz, "Environmental Virtue Ethics: What It Is and What It Needs to Be," in *The Cambridge Companion to Virtue Ethics*, ed. D. C. Russell, Cambridge: Cambridge University Press, 2013, pp. 221–239.

Index

Other volumes in the series of Cambridge Companions

For a list of titles published in the series, please see end of book.

Abelard Edited by Jeffrey E. Brower *and* Kevin Guilfoy

Adorno Edited by Thomas Huhn

Ancient Scepticism Edited by Richard Bett

Anselm Edited by Brian Davies *and* Brian Leftow

Aquinas Edited by Norman Kretzmann *and* Eleonore Stump

Arabic Philosophy Edited by Peter Adamson *and* Richard C. Taylor

Hannah Arendt Edited by Dana Villa

Aristotle Edited by Jonathan Barnes

Aristotle's 'Politics' Edited by Marguerite Deslauriers *and* Paul Destrée

Atheism Edited by Michael Martin

Augustine Edited by Eleonore Stump *and* Norman Kretzmann

Bacon Edited by Markku Peltonen

Berkeley Edited by Kenneth P. Winkler

Boethius Edited by John Marenbon

Brentano Edited by Dale Jacquette

Carnap Edited by Michael Friedman *and* Richard Creath

Constant Edited by Helena Rosenblatt

Critical Theory Edited by Fred Rush

Darwin 2nd edition *Edited by* Jonathan Hodge *and* Gregory Radick

Simone De Beauvoir Edited by Claudia Card

Deleuze Edited by Daniel W. Smith *and* Henry Somers-Hall

Descartes Edited by John Cottingham

Dewey Edited by Molly Cochran

Duns Scotus Edited by Thomas Williams

Early Greek Philosophy Edited by A. A. Long

Early Modern Philosophy Edited by Donald Rutherford

Epicureanism Edited by James Warren

Existentialism Edited by Steven Crowell

Feminism In Philosophy Edited by Miranda Fricker *and* Jennifer Hornsby

Foucault 2nd edition *Edited by* Gary Gutting

Frege Edited by Tom Ricketts *and* Michael Potter

Freud Edited by Jerome Neu

Gadamer Edited by Robert J. Dostal

Galen Edited by R. J. Hankinson

Galileo Edited by Peter Machamer

German Idealism Edited by Karl Ameriks

Greek and Roman Philosophy Edited by David Sedley

Habermas Edited by Stephen K. White

Hayek Edited by Edward Feser

Hegel Edited by Frederick C. Beiser

Hegel and Nineteenth-Century Philosophy Edited by Frederick C. Beiser

Heidegger 2nd edition *Edited by* Charles Guignon

Hobbes Edited by Tom Sorell

Hobbes's 'Leviathan' Edited by Patricia Springborg

Hume 2nd edition *Edited by* David Fate Norton *and* Jacqueline Taylor

Husserl Edited by Barry Smith *and* David Woodruff Smith

William James Edited by Ruth Anna Putnam

Kant Edited by Paul Guyer

Kant and Modern Philosophy Edited by Paul Guyer

Kant's 'Critique Of Pure Reason' Edited by Paul Guyer

Keynes Edited by Roger E. Backhouse and Bradley W. Bateman

Kierkegaard Edited by Alastair Hannay *and* Gordon Daniel Marino

Leibniz Edited by Nicholas Jolley

Levinas Edited by Simon Critchley *and* Robert Bernasconi

Locke Edited by Vere Chappell

Locke's 'Essay Concerning Human Understanding' Edited by Lex Newman

Logical Empiricism Edited by Alan Richardson *and* Thomas Uebel

Maimonides Edited by Kenneth Seeskin

Malebranche Edited by Steven Nadler

Marx Edited by Terrell Carver

Medieval Jewish Philosophy Edited by Daniel H. Frank *and* Oliver Leaman

Medieval Philosophy Edited by A. S. Mcgrade

Merleau-Ponty Edited by Taylor Carman *and* Mark B. N. Hansen

Mill Edited by John Skorupski

Montaigne Edited by Ullrich Langer

Newton Edited by I. Bernard Cohen *and* George E. Smith

Nietzsche Edited by Bernd Magnus *and* Kathleen Higgins

Nozick's 'Anarchy, State And Utopia' Edited by Ralf Bader *and* John Meadowcroft

Oakeshott Edited by Efraim Podoksik

Ockham Edited by Paul Vincent Spade

The 'Origin Of Species' Edited by Michael Ruse *and* Robert J. Richards

Pascal Edited by Nicholas Hammond

Peirce Edited by Cheryl Misak

Philo Edited by Adam Kamesar

The Philosophy Of Biology Edited by David L. Hull *and* Michael Ruse

Piaget Edited by Ulrich MüLler, Jeremy I. M. Carpendale *and* Leslie Smith

Plato Edited by Richard Kraut

Plato's 'Republic' Edited by G. R. F. Ferrari

Plotinus Edited by Lloyd P. Gerson

Pragmatism Edited by Alan Malachowski

Quine Edited by Roger F. Gibson Jr.

Rawls Edited by Samuel Freeman

Thomas Reid Edited by Terence Cuneo *and* René Van Woudenberg

Renaissance Philosophy Edited by James Hankins

Rousseau Edited by Patrick Riley

Bertrand Russell Edited by Nicholas Griffin

Sartre Edited by Christina Howells

Schopenhauer Edited by Christopher Janaway

The Scottish Enlightenment Edited by Alexander Broadie

Adam Smith Edited by Knud Haakonssen

Socrates Edited by Donald Morrison

Spinoza Edited by Don Garrett

Spinoza's 'Ethics' Edited by Olli Koistinen

The Stoics Edited by Brad Inwood

Leo Strauss Edited by Steven B. Smith

Tocqueville Edited by Cheryl B. Welch

Utilitarianism Edited by Ben Eggleston *and* Dale E. Miller

Virtue Ethics Edited by Daniel C. Russell

Wittgenstein Edited by Hans Sluga *and* David Stern

Index