DEREK PARFIT

On What Matters

VOLUME ONE

On What Matters

VOLUME ONE

DEREK PARFIT

Edited and Introduced by
Samuel Scheffler

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP
Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark
of Oxford University Press in the UK and in certain
other countries

Portions of 'On What Matters' by Derek Parfit were delivered as a Tanner Lecture
on Human Values at the University of California, Berkeley, November 2002.
Printed with permission of the Tanner Lectures on Human Values, a Corporation,
University of Utah, Salt Lake City, Utah, USA.

The moral rights of the authors have been asserted
Impression: 1

First published 2011
First published in paperback 2013

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

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Parfit, Derek.
On what matters / Derek Parfit.
p. cm.
Includes bibliographical references and index.
ISBN 978–0–19–957280–9
1. Ethics. I. Title.
BJ1012.P37 2009
170—dc22 2009029662

Typeset by Laserwords Private Limited, Chennai, India
Printed in Great Britain
on acid-free paper by
Clay Ltd., St Ives plc

ISBN 978–0–19–968103–7 (Vol. 1)
978–0–19–968104–4 (Vol. 2)

Cover photograph, by the author, taken through the arch
of the Winter Canal in St. Petersburg.

On What Matters

VOLUME ONE

VOLUME TWO

This page intentionally left blank

Contents

VOLUME ONE

VOLUME TWO

PART FOUR

COMMENTARIES

PART FIVE

RESPONSES

This page intentionally left blank

Introduction

Samuel Scheffler

In this densely argued and deeply original book, Derek Parfit addresses some of the most basic questions in practical philosophy. The book comprises two volumes, each containing three parts. Parfit's central chapters, which make up Parts Two and Three, deal with issues of substantive morality. These chapters descend from a series of three Tanner Lectures that Parfit delivered at the University of California at Berkeley in November of 2002. In Parts One and Six, Parfit addresses issues that were not covered in the Berkeley lectures. Part One is an extended discussion of reasons and rationality, which provides the background for his claims about morality in Parts Two and Three. Part Six takes up the meta-normative questions raised by our use of normative language in making claims both about reasons and about morality.

The three commentators who responded to Parfit's Berkeley Tanner Lectures—Thomas Scanlon, Susan Wolf, and Allen Wood—offer revised versions of their comments in Part Four. In addition, Barbara Herman, who was not a participant in the Berkeley events, contributes a set of comments written specially for inclusion in this book. Parfit replies to all of these comments in Part Five. The exchanges between him and the commentators focus primarily on the chapters deriving from the Berkeley lectures.

In his chapters on morality, Parfit aims to rechart the territory of moral philosophy. Students who take courses in the subject are usually taught that there is a fundamental disagreement between consequentialists, who believe that the rightness of an act is a function solely of its overall consequences, and Kantians, who argue—often with reference to one or another version of "the categorical imperative"—that we have certain duties that we must fulfill whether or not doing so

will produce optimal results in consequentialist terms. Although both consequentialist and Kantian views are acknowledged to admit of many variations and refinements, the division between them is assumed by most philosophers, including most consequentialists and Kantians, to be deep and fundamental.

Parfit's primary aim in Parts Two and Three of this book is to undermine this assumption, and to demonstrate the existence of a startling convergence among positions that we are accustomed to viewing as rivalrous. He begins by engaging in a sustained and searching examination of Kant's own moral philosophy, including his various formulations of the categorical imperative and many of his other central moral ideas as well. Although Kant's ethical writings, especially the *Groundwork of the Metaphysics of Morals*, are among the most widely discussed texts in the history of moral philosophy, Parfit's engagement with these texts yields a wealth of fresh observations and insights.

As is evident from his Preface, Parfit's attitude toward Kant is complex and defies easy summary. He describes him as "the greatest moral philosopher since the ancient Greeks" (235), and says that "in the cascading fireworks of a mere forty pages, Kant gives us more new and fruitful ideas than all the philosophers of several centuries" (183). He quickly adds, however, that "[o]f all the qualities that enable Kant to achieve so much, one is inconsistency" (183). Whereas many commentators explicitly present themselves either as critics of Kant or as defenders of his view, Parfit's approach is different. He treats Kant's texts as a rich fund of claims, arguments, and ideas, all of which deserve to be treated with the same seriousness that one would accord the ideas of a brilliant contemporary, but many of which require clarification or revision, and some of which are simply unworkable. Parfit examines a wide range of these claims, arguments, and ideas, subjecting them to a level of scrutiny that is remarkable for its unwavering focus and analytic intensity. His primary aim is neither to defend Kant nor to criticize him, but rather to determine which of his ideas we can use to make progress in moral philosophy. At the end of the day, it is progress that is Parfit's real goal. As he says in explaining why one of Kant's formulations should be revised, "After learning from the works of great philosophers, we

should try to make some more progress. By standing on the shoulders of giants, we may be able to see further than they could" (300).

Parfit identifies several elements of Kant's thought that he regards as particularly important and that he is prepared to endorse, albeit with some significant revisions and additions. However, he frequently differs from other leading commentators in the way he interprets the content and implications of these ideas. This is perhaps most evident in his treatment of the version of the categorical imperative known as the "Formula of Universal Law." As Parfit observes, this formulation of the categorical imperative has been subject to so many serious objections that many otherwise sympathetic commentators have concluded that it is of little value as an action-guiding principle that can help us to distinguish right from wrong. Many leading Kant scholars have concluded that other formulations of the categorical imperative are richer and more illuminating.

Parfit, by contrast, sees great potential in the Formula of Universal Law. Swimming against the prevailing tide of interpretive opinion, he insists that the FUL "*can* be made to work," and he argues that when "revised in some wholly Kantian ways, this formula is . . . remarkably successful" (294). Indeed, he goes so far as to say that a suitably revised version of this formula "might be what Kant said that he was trying to find: the supreme principle of morality" (342).

The revised version of the Formula of Universal Law that Parfit favors states that "Everyone ought to follow the principles whose universal acceptance everyone could rationally will." With its appeal to a kind of universal choice or agreement, this formulation qualifies as a form of "contractualism," and Parfit refers to it as the "Kantian Contractualist Formula." So interpreted, the Kantian position invites comparison with contemporary versions of contractualism, especially those versions that are themselves of broadly Kantian inspiration. John Rawls's appeal to principles that would be chosen behind a veil of ignorance is one example, though Rawls applied this device almost exclusively to the choice of principles of justice for the basic structure of society. He never followed up on the idea, which he had briefly entertained in *A Theory of Justice*, that the same device might be applied to the choice of moral principles more generally. Parfit nevertheless subjects this idea to severe

criticism, and concludes that it is much less promising as a general account of morality than the version of contractualism developed by Thomas Scanlon.

As Parfit states it, "Scanlon's Formula" holds that "Everyone ought to follow the principles that no one could reasonably reject." Parfit argues that, on some interpretations at least, Scanlonian Contractualism coincides with Kantian Contractualism since, on these interpretations, the principles whose universal acceptance everyone could rationally will turn out to be just the same as the principles that no one could reasonably reject. The possibility of convergence between these two forms of contractualism may not seem terribly surprising, although Parfit and Scanlon disagree about the precise extent of the convergence. What is more surprising is Parfit's assessment of the relations between contractualism and consequentialism.

As I have noted, the opposition between the Kantian and consequentialist positions is usually taken to be deep and fundamental, and the contemporary contractualisms of both Rawls and Scanlon are motivated to a significant degree by the desire to articulate a compelling alternative to consequentialism. Yet Parfit argues that Kantian contractualism actually implies a version of "Rule Consequentialism," which holds that "everyone ought to follow the principles whose universal acceptance would make things go best." The principles whose universal acceptance everyone could rationally will, he maintains, just are these "optimific" rule-consequentialist principles. Accordingly, Kantian Contractualism and Rule Consequentialism can be combined to form a view that he calls Kantian Rule Consequentialism: "Everyone ought to follow the optimific principles, because these are the only principles that everyone could rationally will to be universal laws" (411). Although this position is consequentialist in the content of its claims about the principles that people ought to follow, it is more Kantian than consequentialist in its account of why we should follow these principles. We should follow them because their universal acceptance is something that everyone could rationally will, and not because, as consequentialists would have it, all that ultimately matters is that things should go for the best.

Since Kantian Contractualism implies Rule Consequentialism, and since some versions of Kantian Contractualism coincide with some

versions of Scanlonian Contractualism, versions of all three positions can also be combined. The resulting "Triple Theory" holds that an "act is wrong just when such acts are disallowed by some principle that is optimific, uniquely universally willable, and not reasonably rejectable" (413). The upshot of these various possibilities of convergence, Parfit believes, is that it is a mistake to think that there are deep disagreements among Kantians, contractualists, and consequentialists. Instead, "[t]hese people are climbing the same mountain on different sides" (419).

In developing this central line of argument, Parfit relies heavily on substantive claims about reasons and rationality. The theories he is considering all make claims about the kinds of reasons that people have for wanting and doing various things, and about the conditions under which individuals' actions are reasonable or rational. Accordingly, Parfit's assessment of these theories consists largely in assessing the force of different claims of this sort. But claims about reasons and rationality are scarcely less controversial than claims about right and wrong. Recognizing this, Parfit prefaces his chapters on morality with a detailed exposition and defense of his own views on these topics.

Many philosophers believe that our reasons for action are all provided by our desires. We have most reason to do whatever will best fulfill either our actual desires or the desires that we would have under ideal conditions. Although such desire-based views, which Parfit classifies as "subjective theories," have been profoundly influential, both within and outside of philosophy, Parfit believes that they are deeply misguided, and his criticism of them is withering. Not only do they have wildly implausible implications, he argues, but they are ultimately "built on sand." They imply that our reasons derive their normative force from desires that we have no reason to have; but such desires, he argues, cannot themselves be said to give us reasons. In the end, then, the real implication of desire-based views is that we have no reasons for action at all and, more fundamentally, that nothing really matters, in the sense that we have no reason to care about any of the things we do care about.

Rejecting these "bleak" views, Parfit argues that we should instead accept an objective, value-based theory, according to which reasons for action are provided by the values that those acts would realize

or fulfill (or, as he puts it, by the facts that make certain things worth doing for their own sake or make certain outcomes good or bad). Understood in this way, judgments about reasons are more fundamental than judgments about rationality, for we are rational, in Parfit's view, when we respond to reasons or apparent reasons, and our acts are rational when, if our beliefs were true, we would be doing what we had good reasons to do. This contrasts with a number of popular accounts of practical rationality, such as those that identify it with the maximization of expected utility, for example, or those that interpret practical *ir*rationality as a form of inconsistency.

As Thomas Scanlon observes in his contribution, the idea that reasons have priority over rationality also conflicts with Kant's views. For Kant, both the authority and the content of the categorical imperative are to be understood with reference to the requirements of rational agency rather than to some independent conception of the reasons that people have. As Scanlon describes the Kantian view, which he calls "Kantian constructivism about reasons": "Claims about reasons (more exactly, about what a person must see as reasons) must be grounded in claims about rational agency, claims about what attitudes a person can take, consistent with seeing herself as a rational agent. Justification never runs in the other direction, from claims about reasons to claims about what rationality requires" (Volume Two, *118*).*

Parfit, like Scanlon, rejects Kantian constructivism about reasons and, as Scanlon points out, all of the moral theories whose convergence Parfit seeks to demonstrate are framed in such a way as to "appeal to an idea of 'what one can rationally will' that presupposes an independently understandable notion of the reasons that a person has and their relative strength" (*118*). This distinguishes these theories from Kant's own views and also from the views of some prominent contemporary Kantians, such as Christine Korsgaard. As Parfit acknowledges, his reliance on a primitive and "indefinable" notion of "reasons," and his concomitant commitment to the existence of irreducibly normative truths, both about reasons and about morality, makes his view a version of what Korsgaard has called "dogmatic rationalism." As such, it would be resisted not only by Kantian constructivists like Korsgaard but also

* Page numbers in italics refer to Volume Two.

by proponents of some very different meta-ethical outlooks, such as various forms of naturalism and non-cognitivism.

In Part Six, therefore, Parfit undertakes to explain and defend his conception of normativity. He endorses a view that he refers to as "Non-Metaphysical Non-Naturalist Cognitivism," which appeals to certain intuitive beliefs we are said to have about irreducibly normative truths. This view is not Platonistic in the sense of making claims about some supposed non-spatio-temporal portion of reality. Nor is its reliance on intuitions meant to suggest that normative facts are apprehended via a mental faculty that is analogous to sense perception. We do not detect the presence of normative properties like rightness or rationality as a result of being causally affected by them. Instead, we understand normative truths in something like the way we understand mathematical or logical truths. Indeed, Parfit argues, mathematical and logical reasoning themselves involve recognizing and responding to normative truths about what we have reason to believe. For example, we recognize that the truth of *p* and *if p then q* gives us conclusive reason to believe that *q*. Just as there are truths about what we have reason to believe, Parfit insists, so too there are truths about what we have reason to do.

Parfit realizes, of course, that many philosophers do not accept the existence of irreducibly normative truths in his sense. Nihilists and error theorists hold that all normative claims are false. Naturalists hold that normative facts can be reduced to natural facts. Non-cognitivists hold that normative claims, despite their importance in human life, do not function as statements of fact at all. Parfit discusses and criticizes many influential versions of such positions, including the views of Simon Blackburn, Richard Brandt, Allan Gibbard, Richard Hare, John Mackie, and Bernard Williams. None of these views, he argues, can adequately account for the normative dimension of our thought; on all such views, normativity proves to be illusory. It simply disappears. In effect, Parfit appears to believe that all such views tend toward nihilism, and that nihilism is the only genuine alternative to the recognition of irreducibly normative truths. Nor is he persuaded by Korsgaard's Kantian objections to "realism" about normativity. Contrary to what she maintains, he asserts, normativity does not have its source in the

will, but instead consists in the existence of irreducibly normative truths about what we have reason to believe, to want, and to do.

As will be apparent, Parfit's aims in his discussions of reasons and normativity are very different from those he pursues in discussing substantive moral theories. In the moral case, his aim is to demonstrate that certain putatively opposing theories may actually converge, so that apparent disagreement among them evaporates. But in his discussion of different views about reasons and normativity, a convergence among rival theories is not on the agenda. Instead, he argues that a value-based theory of reasons should be accepted and that desire-based theories should be rejected. Similarly, his form of Cognitivism should be accepted in preference to all forms of Naturalism and Non-Cognitivism. Parfit is clearly troubled by substantive moral disagreement, for he thinks it threatens to undermine our conviction that there is such a thing as moral truth. That is why he is so strongly driven to demonstrate the possibility of convergence among rival moral theories. Although he is also troubled by meta-ethical or meta-normative disagreement, his response to it is different. Here he simply attempts to determine which of the contending positions is correct. Yet to the extent that the substantive moral theories whose convergence Parfit seeks to demonstrate all presuppose his views about reasons and normativity, the frankly contested character of those views may call into question the significance of the convergence he describes at the substantive moral level. Those who reject value-based theories of reasons, and those who accept one or another form of naturalism or non-cognitivism or constructivism, may be unmoved by a moral consensus that depends on accepting the very meta-ethical views that they reject. So one challenge for Parfit is to demonstrate that the significance of the convergence for which he argues is not undermined by its dependence on claims, such as those concerning reasons and normativity, about which there is no convergence. Although Parfit does not directly address this challenge, he does argue that those who have rejected the views about reasons and normativity that he favors have not always fully understood them. And he expresses the hope that, once the relevant misunderstandings have been cleared away, many more philosophers will eventually come to accept those views. If this is correct, then even though the competing theories of reasons and of

normativity do not themselves converge, there may be reason to hope for much greater convergence in the assessments that philosophers give of them. Of course, this suggestion is itself likely to be controversial.

There are many other questions that can and will be raised about Parfit's subtle and intricate arguments. One issue, different aspects of which are discussed by each of the four commentators, concerns the extent to which the views whose convergence Parfit seeks to demonstrate are authentic versions of more familiar moral views. To what extent is Kantian Contractualism really Kantian? We have already seen that, in its account of the relation between rationality and reasons, the view appears to be more Parfit's than Kant's. Similar questions can be raised about the other ostensibly convergent positions. To what extent does Scanlonian Contractualism reflect Scanlon's own views? And what is the relation between Parfit's version of Rule Consequentialism and other consequentialist formulations?

The issue is a tricky one. As Scanlon notes, Parfit is forthright about his willingness, in developing a "Kantian" position, to depart from Kant's actual views whenever he thinks he can improve upon them. As Parfit says, "We are asking whether Kant's formulas can help us to decide which acts are wrong, and help to explain why these acts are wrong. If we can revise these formulas in ways that are clearly needed, we are developing a Kantian moral theory" (298). In his reply to Scanlon, he is similarly explicit about the fact that his argument for the convergence of Kantian Rule Consequentialism and Scanlonian Contractualism "does not apply to the view stated in Scanlon's book" (*244*), but rather to a version of that view that has been revised in ways that Parfit takes to strengthen it.

This unapologetic revisionism carries with it two risks for Parfit. The first, which Scanlon mentions, is that the degree to which any convergence he can demonstrate will seem surprising and significant may depend on how close the convergent theories are to the eponymous ancestors from which they descend. The more they have been revised in ways that depart from their original formulations, the less surprising and significant their convergence may seem. The second risk is that, in revising the original theories to bring them closer to one another, valuable elements of the original theories may be excluded.

Susan Wolf appears to harbor doubts of both of these kinds about Parfit's claims of convergence. Of Parfit's ambition to reconcile the Kantian, consequentialist, and contractualist traditions, she writes: "[I]nsofar as the remarks quoted above are meant to suggest that the values these different traditions emphasize can be interpreted and ordered in such a way as to eliminate the tensions among them, or that it would be in the spirit of these traditions' greatest exponents to accept revisions and qualifications to their stated views that would ultimately reconcile them with their opponents, Parfit departs from the explicit positions of any of the philosophers whose work he discusses, in a way that seems to me both interpretively implausible and normatively regrettable" (*32*). Wolf's view is that the Kantian, consequentialist, and contractualist traditions embody divergent evaluative perspectives, each of which has something important to contribute but which are in genuine tension with one another. These tensions reflect broader tensions within our moral thought itself. As such, she believes, they are ineliminable and not to be regretted. Any unified principle of the kind Parfit seeks will perforce be a matter of compromise rather than complete convergence, and any such principle will inevitably leave out something of value. Wolf presses this last point with special reference to Parfit's version of Kantianism, which, she argues, scants the importance of autonomy in Kant's own moral philosophy.

Barbara Herman too believes that Parfit's position departs from Kant's in fundamental ways. However, while Wolf expresses doubts about the very idea that morality rests on a unified principle of the kind that Parfit seeks, Herman is sympathetic to Kant's own unified account and believes that Parfit's theory is an unstable mixture of disparate elements. More specifically, she argues that Parfit employs a "hybrid" methodology that incorporates some Kantian features but nevertheless has "a strongly consequentialist cast" (*81*). Although Parfit's intention is to preserve what is most persuasive in Kant's view while avoiding some of the apparently unwelcome implications of that view, Herman believes that there is such a deep "mismatch" between the Kantian and consequentialist methodologies that the attempt to combine them inevitably distorts Kant's own account and obscures what is most appealing about it. In the first portion of her comments, she identifies

several elements of Parfit's methodology that she regards as deeply consequentialist in character, and she gives illustrations of the resulting methodological divide that she sees between Parfit and Kant. Perhaps the most basic difference is this: whereas Parfit appeals to various nonmoral goods to determine what people could rationally will and so to fix the content of morality itself, Kant, Herman says, seeks to establish a place for nonmoral goods within an independently established moral framework. In the remainder of her commentary, she attempts to demonstrate that this "unified" Kantian approach, properly developed, has the resources to accommodate some of the most important moral intuitions—such as those concerning permissible lies—that Kant has seemed to neglect. If this is correct, then much of the motivation for a hybrid moral methodology disappears. In his reply, Parfit does not directly engage with Herman's thoughtful attempt to develop the unified Kantian view in this way. However, he disputes her assessment of the "mismatch" between his methodology and Kant's. Most of the ostensibly consequentialist aspects of his method that she cites, he maintains, are also features of Kant's view. And although he does propose revisions in Kant's Formula of Universal Law, some of these revisions are fully in the spirit of the Kantian view, while others are necessary to avoid straightforward mistakes. The upshot, Parfit believes, is that the gap between his own position and Kant's is far narrower, and far shallower, than Herman asserts.

Like Herman, Allen Wood also argues that Parfit's philosophical methodology departs from Kant's in important ways, although he focuses on different aspects of Parfit's approach than Herman does. Wood believes that Parfit employs a method originated by Sidgwick, which sets itself the goal of providing a "scientific" ethics. The idea is to systematize our commonsense moral opinions, correcting them when necessary, with the aim of arriving at a precise set of principles that can be used algorithmically to yield a determinate moral verdict about how one should act in any conceivable situation. Wood believes that such otherwise diverse philosophers as Kant, Bentham, and Mill employ a very different method, which he himself regards as preferable to the one he ascribes to Sidgwick and Parfit. This alternative method begins not with commonsense intuitions but rather with a fundamental principle

that serves to articulate some basic value. General moral rules or duties are then derived non-deductively from the fundamental principle. These rules or duties represent an attempt to interpret the implications of the fundamental value in the conditions of human life. The rules or duties themselves admit of exceptions and require interpretation, and their application to particular cases calls for the exercise of judgment and cannot be codified in precise rules or principles. So, on the one hand, the Kantian method as Wood understands it gives less weight than the Sidgwickian method to commonsense moral intuitions; but, on the other hand, it regards as "hopeless" the aim of constructing a "scientific" ethics that can provide an algorithm for moral decision-making.

Wood believes—though Parfit's reply suggests that he would not accept this diagnosis—that the difference of method just described underlies some disagreements between Parfit and him concerning the proper interpretation of Kant's Formula of Humanity. He thinks it also underlies their sharply divergent attitudes toward one familiar type of philosophical argument. This type of argument uses our intuitive reactions to stylized and sometimes complex hypothetical examples to test candidate moral principles. Wood refers to all such examples as "trolley problems," whether or not they involve actual trolleys, in mock *hommage* to the famous case first introduced into the philosophical literature by Philippa Foot. Parfit makes frequent use of such examples in constructing his arguments. For instance, his argument for the convergence of Kantian Contractualism and Rule Consequentialism turns crucially on some claims about what a person could rationally agree to in situations where one course of action would impose a burden on the person himself and the only alternative would impose burdens on others. Parfit illustrates and defends these claims with reference to a series of hypothetical examples involving burdens of different sizes and types imposed in a range of different hypothetical circumstances. He seeks to marshal our intuitive responses in these cases to show (1) that each person could rationally will the universal acceptance of the consequentially optimific principles, even when those principles would impose some burden on the person himself, and (2) that there are no other principles whose universal acceptance everyone could rationally choose. Parfit evidently believes that the use of hypothetical

examples can help to clarify the issues that are at stake in complex moral choices and enable us to make progress in moral argument. Wood, by contrast, regards "trolley problems" as "worse than useless for moral philosophy" (*68*), and the majority of his essay is given over to an extended critique of the ways in which reliance on such problems leads moral philosophers astray.

To the extent that other people share Wood's reservations about appealing to hypothetical examples in moral philosophy, Parfit's extensive reliance on such examples may be a source of resistance to his arguments. Of course, even those who do not endorse Wood's radical rejection of all such appeals may find themselves disagreeing with Parfit's reactions to some of the specific examples he discusses, although Parfit anticipates many potential disagreements and exhibits great resourcefulness in attempting to defuse them. Yet Parfit himself points out that our reactions to some of these cases may depend, for example, on whether we accept a desire-based or a value-based theory of reasons. Since he hopes to use our reactions to support his claim of convergence among different moral theories, this kind of variation represents one way in which disagreements about reasons and rationality, like meta-ethical disagreements about the nature of normative judgment, threaten to destabilize the moral consensus that Parfit aims to establish. As I have already said, Parfit's response to this threat is not to look for convergence among the rival meta-ethical theories or theories of reasons and rationality themselves. Instead, he argues that there are decisive reasons for rejecting the alternatives to Non-Metaphysical Non-Naturalist Cognitivism and the value-based theory of reasons, and he pins his hopes for convergence on the possibility that philosophers will eventually come to accept the cognitivist and value-based positions that he favors. This is a different way of eliminating or at least taking the sting out of disagreement: by demonstrating that there is only one position that we can reasonably accept.

The drive to eliminate disagreement—whether by establishing theoretical convergence or through a decisive demonstration of the inadequacy of competing views—is a defining feature of Parfit's work. It is sometimes marked by a sense of urgency. One place where this emerges is in his reply to Susan Wolf. Wolf takes Parfit to be trying

to show "that there is a single true morality, crystallized in a single supreme principle which these different traditions may be seen to be groping towards, each in their own separate and imperfect ways" (*32*). She herself says, by contrast, that "it would not be a moral tragedy if it turned out" that morality did not have such a unifying principle (*33*). In response, Parfit agrees that it would not be a tragedy if there were no single supreme principle. But, he adds, "it *would* be a tragedy if there was no *single true morality*." He adds: "if we cannot resolve our disagreements, that would give us reasons to doubt that there are *any* true principles. There might be nothing that morality *turns out to be*, since morality might be an illusion." (*151*). It is, perhaps, the spectre of this "bleak" possibility, and the even bleaker possibility that, as Parfit worries, nothing at all may matter, that is responsible for the sense of urgency with which he pursues the elimination of disagreement. Whether or not one shares his assessment of the threat posed by deep disagreement, one cannot fail to be impressed by the extraordinary ingenuity and the sheer intellectual intensity with which he pursues his goal. His rich and challenging discussion, helpfully illuminated by his exchanges with Barbara Herman, Thomas Scanlon, Susan Wolf, and Allen Wood, casts familiar debates in a fresh and unfamiliar light, and opens up many fruitful new lines of inquiry for philosophers to investigate. Nobody who is interested in the theory of morality, rationality, or normativity will want to ignore this brilliant, provocative, and tenaciously argued book.

Preface

Since this book contains summaries, I shall say little about its contents here. Though the book is long, there are some shorter books within it. Nothing important in Part Three depends on Part Two, so you might read only Parts One and Three. If you are mainly interested in ethics, you might read only Chapters 6 to 17. If you are mainly interested in reasons, rationality, and meta-ethics, you might read only Parts One and Six.

While describing how he came to write his great, drab book *The Methods of Ethics*, Sidgwick remarks that he had 'two masters': Kant and Mill. My two masters are Sidgwick and Kant.

Kant is the greatest moral philosopher since the ancient Greeks. Sidgwick's *Methods* is, I believe, the best book on ethics ever written. There are some books that are greater achievements, such as Plato's *Republic* and Aristotle's *Ethics*. But Sidgwick's book contains the largest number of true and important claims. It is not surprising that, though a less great philosopher than Plato, Aristotle, Hume, and Kant, Sidgwick could write a better book. Sidgwick lived later. Unlike later poets or playwrights, who have no advantages over Homer or Shakespeare, later philosophers do have advantages, since philosophy makes progress.

Sidgwick and Kant both have weaknesses and flaws. Sidgwick is sometimes boring, for example, and Kant is sometimes maddening. I hope that by admitting these weaknesses, and saying why we should not be disappointed or deterred by them, I may persuade some people to read, or re-read, Sidgwick's *Methods* and some of Kant's books.

Kant and Sidgwick are a wonderfully contrasting pair. Discussing their own achievements, for example, Kant writes:

> . . . the critical philosophy must remain confident of its irresistible propensity to satisfy the theoretical as well as the moral, practical purposes of reason, confident that no change of opinions, no touching up or reconstruction into some other form, is in store for it; the system of the *Critique* rests on a fully secured foundation, established forever; it will prove to be indispensable too for the noblest ends of mankind in all future ages;

Sidgwick writes:

> The book solves nothing, but may clear up the ideas of one or two people, a little.

Kant is very original, makes some sublime claims, and is excitingly intense. Sidgwick knew that he lacked these qualities. 'I like criticizing myself', he writes to a friend, 'and have formulated the following on it:

> *Pro*: Always thoughtful, often subtle: generally sensible and impartial: approaches the subject from the right point of view.
>
> *Con*: Inconsequent, ill-arranged: stiff and ponderous in style, nothing really striking or original in the arguments.'

Sidgwick also refers to his 'one damning defect of longwinded & difficult dullness'.

This last phrase is too severe. Though Sidgwick's book is long, and some of its chapters can now be ignored, it is not longwinded. Sidgwick seldom repeats himself, and he makes many important points concisely, and only once. Nor is Sidgwick's book difficult. Some of his claims and arguments are complicated, but they are nearly all clearly written.

Sidgwick's dullness needs more discussion. Whitehead was so bored by Sidgwick's *Methods* that he never looked at another book on ethics.

But after reading a collection of Sidgwick's memoirs and letters, Keynes remarked, 'I have never found so dull a book so absorbing'. It is worth quoting from this book. Discussing the Church of England, Sidgwick writes:

> At Cambridge I get into the way of regarding it as something that once was alive and growing, but now exists merely because it is a pillar or buttress of uncertain value in a complicated edifice that no one wants just now to take to pieces. Here however, I feel rather as if I were contemplating a big fish out of water, propelling itself smoothly and gaily over the high road.

Here are two more passages:

> There is no doubt that men in England fall in love chiefly in abnormal periods: when on a reading party, or at the seaside, or at a foreign hotel, or at Christmas, or any other occasion when something, either external circumstances or any dominant emotion, thaws the eternal ice. The misfortune is that if these casual thaws do not last long enough, all the advantage gained is lost; two lines of life that causally intersected diverge perhaps for ever, and the frost sets in with redoubled force.

> I am bearing the burden of humanity in the lap of luxury, and in consequence not bearing it well. After all, Pascal was practically right: if one is to embrace infinite doubt, if it is to come into our bowels like water, and like oil into our bones, it ought to be upon sackcloth and ashes and in a bare cell, and not amid '47 port and the silvery talk of W. G. Clark. When I go to my rooms I feel strange, ghastly, that is why I write to you. But there again—if one allows this consciousness 'the time is short' to grow and get too strong, it seems to fold up all life into a feverish moment.
>
> The world shall feel my impulse or I die.

> Think of all the second-rate men who have said this and died—and—Who cares?
>
> Butterflies may dread extinction.
>
> This is a strange mood for me. But at Trumpington today I brushed away a spider's life and said 'This is sentience.' What am I more than elaborate sentience?

Sidgwick could be amusing, and his conversation was described as 'like the sparkling of a brook whose ripples seem to give out sunshine'. But the first edition of the *Methods* contains only a few jokes, some of which Sidgwick later removed. Much of the book, however, is well-written. For example:

> to suppose . . . that the ideal of 'obeying oneself alone' can be even approximately realized by Representative Democracy is even more patently absurd. For a representative assembly is normally chosen only by a part of the nation, and each law is approved by only a part of the assembly: and it would be ridiculous to say that a man has assented to a law passed by a mere majority of an assembly *against* one member of which he has voted.

More soberly:

> . . . 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 foredoomed to inevitable failure.

This magnificently sombre claim has some of the intensity of Kant, as does another passage that is about Kant:

> I cannot fall back on the resource of thinking myself under a moral necessity to regard all my duties *as if they were* commandments of God, although not entitled to hold speculatively that any such Supreme Being really exists. I am so far from feeling bound to believe for purposes of practice what I see no ground for holding as a speculative truth, that I

> cannot even conceive the state of mind which these words seem to describe, except as a momentary half-witted irrationality, committed in a violent access of philosophic despair.

Many fine passages are too long to quote in full. One such passage ends:

> . . . the 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 . . . the discord between the rhythms of his own life and of that larger life of which his own is but an insignificant fraction.

Another passage ends:

> . . . even a man who said 'Evil be thou my good' and acted accordingly might have only an obscured consciousness of the awful irrationality of his action — obscured by a fallacious imagination that his only chance of being in any way admirable, at the point of which he has now reached in his downward course, must lie in candid and consistent wickedness.

Sidgwick warned his friends that, because his book attempts to achieve 'precision of thought', it 'cannot fail to be somewhat dry and repellent'. But this precision is often finely expressed. Discussing friendship, for example, Sidgwick describes

> the sympathy that is not quite admiration with which Common Sense regards all close and strong affections; and the regret that is not quite disapproval with which it contemplates their decay.

Many sentences, though dry, have an ironical edge or twist. For example:

> It may be said that a child owes gratitude to the authors of its existence. But life alone, apart from any provision for making life happy, seems a boon of doubtful value, and one that scarcely excites gratitude when it was not conferred from any regard for the recipient.

> . . . there seems to be no justice in making A happier than B, merely because circumstances beyond his control have first made him better.

> Thus the Utilitarian conclusion, carefully stated, would seem to be this: that the opinion that secrecy may render an action right which would not otherwise be so should itself be kept comparatively secret; and similarly it seems expedient that the doctrine that esoteric morality is expedient should itself be kept esoteric.

> . . . really penetrating criticism, especially in ethics, requires a patient effort of sympathy which Mr Bradley has never learned to make, and a tranquillity of temper which he seems incapable of maintaining.

> [The book] seems smashing, but he loses by being over-controversial. There should be at least an affectation of fairness in a damaging attack of this kind.

Sidgwick's irony can make him seem stuffy, when in fact he is being subversive. Bernard Williams had been misled, for example, when he wrote that Sidgwick's discussions of sexual morality, though sometimes mildly adventurous, 'make fairly uncritical use of a notion of purity'. Sidgwick does ask 'What, then, is the conduct that Purity forbids?' But if we read him carefully, we find that his answer is: Nothing. In a book published in England in 1874, it was more than mildly adventurous to argue, though in guarded terms, that there is no moral objection to indulging in sexual pleasure for its own sake.

When people find Sidgwick dull, they are often responding not to Sidgwick's style, but to one of his greatest philosophical merits. Sidgwick describes this merit well, writing in his journal:

> Have been reading Comte and Spencer, with all my old admiration for their intellectual force and industry and more than my old amazement at their fatuous self-confidence. It does not seem to me that either of them knows what self-criticism means. I wonder if this is a defect inseparable from their excellences. Certainly I find my own self-criticism an obstacle to energetic and spirited work: but on the other hand I feel that whatever value my work has is due to it.

Sidgwick was unusually good at seeing the force of objections to his views. After hearing Sidgwick defend a paper, William James remarked:

> Sidgwick displayed that reflective candour that can at times be so irritating. A man has no right to be so fair to his opponents.

Discussing an opponent's book, for example, Sidgwick writes:

> I shall praise it as much as I can . . . it is by an author of fine qualities . . . But yet—he seems to me altogether out of it: I can scarcely treat his theory with proper respect. No doubt I seem so to him: and are we not both right? The book makes me rather depressed about ethics.

These virtues can make Sidgwick hard to read. One problem is that, as C. D. Broad explains, Sidgwick

> incessantly refines, qualifies, raises objections, answers them, and then finds further objections to the answer. Each of these objections, rebuttals, rejoinders, and surrejoinders is in itself admirable, and does infinite credit to the acuteness and candour of the author. But the reader is apt to become impatient; to lose the thread of the argument; and to rise from his desk finding that he has read a great deal with constant admiration and now remembers little or nothing.

Our first reading of the *Methods* is, in a way, the worst, since there is little that is striking or inspiring. But every time we re-read this book, we notice some new good points that we had earlier overlooked. That is what I, at least, have found.

Criticizing himself again, Sidgwick writes:

> I am not an original man: and I think less of my own thoughts every day.

This remark is also too severe. Sidgwick is in several ways original. But that is not what makes him great. Other philosophers, like Kant and Hume, are more original, and more brilliant. These philosophers are like Newton and Einstein: geniuses of the clearest kind. Sidgwick is more like Darwin. He had what has been called 'good sense intensified almost to the point of genius'. In the *Methods*, as Broad claims, 'almost all the main problems of ethics are discussed with extreme acuteness'. And Sidgwick gets very many things right. He gives the best critical accounts of three of the main subjects in ancient and modern ethics: hedonism, egoism, and consequentialism. And in the longest of his book's four parts, he also gives the best critical account of pluralistic non-consequentialist common sense morality. Though Sidgwick makes mistakes, some of which I mention in a note, he does not, I believe, make many. These facts make Sidgwick's *Methods* the book that it would be best for everyone interested in ethics to read, remember, and be able to assume that others have read.

My debts to Sidgwick are easy to describe. Of my reasons for becoming a graduate student in philosophy, one was the fact that, in wondering how to spend my life, I found it hard to decide what really matters. I knew that philosophers tried to answer this question, and to become wise. It was disappointing to find that most of the philosophers who taught me, or whom I was told to read, believed that the question 'What matters?' couldn't have a true answer, or didn't even make sense. But I bought a second-hand copy of Sidgwick's book, and I found that he at least believed that some things matter. And it was from Sidgwick that I learnt most about the other questions that moral philosophers should ask, and about some of the answers.

I turn now to my other master, Kant. When I first read Kant's *Groundwork* in the 1960s, I found this book fascinating but obscure. When I re-read this book thirty years later, and most of Kant's other books, I became unexpectedly obsessed with Kant's ethics. For the next two or three years, I thought about little else.

It seems worth confessing that, though my obsession with Kant gave me great energy, this energy was, to start with, almost entirely negative. I didn't doubt Kant's genius. But like many other people, I found myself deeply opposed both to some of Kant's main claims, and to his way of doing philosophy. By mentioning what made me so opposed to Kant, and saying how my attitude has changed, I may perhaps persuade some other people not to ignore Kant, as I nearly did.

Though Kant has some important qualities that Sidgwick lacks, Kant also lacks some important qualities that Sidgwick has. Sidgwick writes clearly, is on the whole consistent, and makes few mistakes. These things cannot be claimed of Kant.

Unlike our first reading of Sidgwick's *Methods*, our first reading of Kant's *Groundwork* is, in some ways, the best. There are some striking and inspiring claims, and we are not worried by what we can't understand. But when we re-read the *Groundwork*, many of us become discouraged, and give up. We decide that Kant, though he may be a great philosopher, is not for us.

The first problem is Kant's style. It is Kant who made really bad writing philosophically acceptable. We can no longer point to some atrocious sentence by someone else, and say 'How can it be worth reading anyone who writes like that?' The answer could always be 'What about Kant?'

There are deeper problems. When I became obsessed with Kant, I tried to restate more clearly some of Kant's main claims and arguments, and found this task very frustrating. I couldn't fit Kant's claims together in a coherent view, and many of Kant's arguments seemed to be obviously invalid or unsound. It would have helped me to know that even some of Kant's greatest admirers have similar feelings. Onora O'Neill, for example, calls the *Groundwork* 'the most exasperating' of Kant's books.

It would also have helped me to know that Kant did not have a single, coherent theory. When we ask whether Kant accepts or rejects some claim, the answer is often 'Both'. As Kemp Smith writes, 'citation of single passages is quite inconclusive'. For example, though Kant writes that 'a human being's duty at each instant is to do all the good in his power', he is not really, as this claim implies, an Act Consequentialist. Rawls remarks that, when he tried to understand Kant's texts, 'I assumed there were never plain mistakes, not ones that mattered anyway'. But there must be mistakes, since Kant makes many conflicting claims, and such claims cannot all be true. As Kemp Smith points out, Kant often 'flatly contradicts himself' and 'there is hardly a technical term which is not employed by him in a variety of different and conflicting senses. He is the least exact of the great thinkers.' (To avoid provoking Hegelians, we should perhaps say 'one of the least exact'.)

'Consistency', Kant writes, 'is a philosopher's greatest duty.' That is not true. Originality and clarity are at least as important. And Kant's greatness chiefly consists in his having many original and fruitful ideas. If Kant had always been consistent, he could not have had all these ideas.

When I first re-read Kant, what I found most irritating was not Kant's obscurities and inconsistencies, but a particular kind of overblown, false rhetoric. For example, Kant writes:

> If we look back upon all previous efforts that have ever been made to discover the principle of morality, we need not wonder why all of them had to fail. It was seen that the human being is bound to laws by his duty; but it never occurred to them that he is subject only to laws given by himself but still universal and that he is obligated only to act in conformity with his own will . . .

I didn't mind the exaggeration in the first sentence here. We can switch the volume down, turning 'all of them had to' into 'some of them did'. But since I knew that Kant believed in a Categorical Imperative, I was surprised by Kant's second sentence. I asked a Kantian, 'Does this mean that, if I don't give myself Kant's Imperative as a law, I am not subject

to it?' 'No,' I was told, 'you have to give yourself a law, and there's only one law.' This reply was maddening, like the propaganda of the so-called 'People's Democracies' of the old Soviet bloc, in which voting was compulsory and there was only one candidate. And when I said 'But I haven't given myself Kant's Imperative as a law', I was told 'Yes you have'. This reply was even worse. My irritation at such claims may have left some traces in this book.

As I have said, however, that irritation has gone. Now that I have read Kant's other works, I am aware of the passions that led Kant to make his most outrageous claims. When he is calmer, he makes other, better claims. For example, Kant is reported to have said:

> Suicide is the most abominable of the crimes that inspire horror and hatred . . . he who so utterly fails to respect his life . . . can in no way be restrained from the most appalling vices . . .

But he also said:

> In the Stoic's principle concerning suicide there lay much sublimity of soul: that we may depart from life as we leave a smoky room.

Some of Kant's impassioned arguments, moreover, have great charm. When condemning suicide, Kant said:

> If freedom is the condition of life, it cannot be employed to abolish life . . . Life is supposedly being used to bring about lifelessness, but that is a self-contradiction.

It is the word 'supposedly' that is so endearing here. Suicide involves a contradiction, one commentator suggests, because it is we, on Kant's view, who confer value on our ends. If we kill ourselves to avoid suffering, we

> cut off the source of the goodness of this end—it is no longer really an end at all, and it is no longer rational to pursue it.

This conclusion arrives too late.

For another example, consider Kant's claim that, if we tell some lie 'even to achieve some really good end', we 'violate the dignity of humanity in our own person' and make ourselves a 'mere deceptive appearance of a human being', who has 'even less worth than if he were a mere thing'. We should ignore such outbursts. On the very next page Kant suggests that, if we are asked by an author whether we like his work, we may be permitted to say what he expects.

Kant is sometimes thought of as a cold, dry, rationalist. But he is really an emotional extremist. As Sidgwick writes, 'Oh, how I sympathize with Kant! with his passionate yearning for synthesis and condemned by his reason to criticism . . . ' Kant seldom uses words like 'most', 'many', 'several', or 'some', preferring to write only 'all' or 'none'. Kant uses 'good', he says, to mean 'practically necessary'. And he seldom uses the concept of a reason: a fact that merely *counts in favour* of some act, since his preferred normative concepts are *required*, *permitted*, and *forbidden*. Temperamentally, I am an extremist too, who has to struggle to be more like Sidgwick.

Oxford University once had a useful marking grade: *Alpha Gamma*. As everyone should agree, Kant's books are pure Alpha Gamma, containing nothing that is *Beta*, or mediocre. Our disagreement should be only about how much of what Kant wrote is Alpha, and how much is Gamma. And if we have found what is Alpha, we can ignore what is Gamma.

Some of Kant's views are, I believe, too close to Hume's. Kant is a more dangerous Anti-Rationalist because, unlike Hume, he seems to be exalting what he calls *Pure Reason*. And Kant's influence has been, I believe, in some other ways bad. But he is very great, and his influence has been, in other and less obvious ways, good. Though Kant makes many claims that are false, and many of his arguments fail, he also gives us some profound truths. Like Sidgwick, I sometimes find him 'quite a revelation'. Kant's books are very thought-provoking. As Rawls writes, 'Part of the wonderful character of the works we study is the depth and variety of ways they can speak to us.'

In this book I try to say something about most of Kant's formulations of his supreme principle of morality. That is why I wrote much of Part Two, though the book's main arguments are in Parts One, Three, and Six. But except in a few sections, which are mostly in Part Two or Appendices F to I, I do not discuss the details of Kant's views.

I turn now to the other people from whom I have learned most. When I was young, most philosophers believed that there could not be any normative truths. So did most economists, other social scientists, and much of the wider Western world. Well-educated non-religious people took for granted the distinction between facts, which are objective, and mere values. Little has changed. When some economist recently claimed that his proposals involved no value judgments, someone else said 'Yes they do. You assume that we ought to do what would be better for some people and worse for no one.' 'That's not a *value judgment*,' this economist replied, 'Everyone accepts it'.

As well as finding, in the long-dead Sidgwick, someone who had greater hopes for practical and moral philosophy, I was encouraged to find some living philosophers who had such hopes. I was encouraged most by Thomas Nagel, and in particular by Nagel's claims about reasons, and about irreducibly normative truths. I have also learnt a great deal from Tim Scanlon. I often cannot remember whether some thought was mine or his. I dedicate this book to these two people.

I am grateful to Christine Korsgaard, whose impressive books led me to reread Kant, and whose critique of what she calls 'dogmatic rationalism' helped to rouse me from my undogmatic slumbers. I have also learnt much (even if not enough) from the remarkable recent series of other books and articles on or inspired by Kant, by such writers as Henry Allison, Marcia Baron, David Cummiskey, Richard Dean, Jeffrey Edwards, Stephen Engstrom, Paul Guyer, Barbara Herman, Thomas Hill, Samuel Kerstein, Patricia Kitcher, Onora O'Neill, Thomas Pogge, Andrews Reath, Jerome Schneewind, David Sussman, Roger Sullivan, and Allen Wood.

I have been greatly helped by many other people, who gave me comments on early drafts. Since it would be impossible to describe in a few pages the many ways in which I have been helped, I can only express my great gratitude to these people.

Of those who gave me comments on all parts of this book, I owe most to Robert Audi, Selim Berker, Talbot Brewer, John Broome, Ruth Chang (to whom I dedicate Chapter 16), Eugene Chislenko, Jerry Cohen, Garrett Cullity, Jonathan Dancy, David Enoch, William Fitzpatrick, Shelly Kagan, Guy Kahane, Niko Kolodny, Michael Otsuka, Ingmar Persson, Jacob Ross, Kieran Setiya, and Larry Temkin. Some parts of this book were jointly written with these people.

I was also greatly helped by Marcello Antosh, Richard Arneson, Rüdiger Bittner, Mary Coleman, Roger Crisp, Stephen Darwall, Harry Gensler, Reto Givel, Elizabeth Harman, Brad Hooker, Frances Kamm, Joseph Mendola, Jefferson McMahan, Liam Murphy, Leonard Katz, Robert Myers, Martin O'Neill, Douglas Portmore, Stuart Rachels, Peter Railton, Karl Schafer, Samuel Scheffler, Michael Slote, Saul Smilansky, Jussi Suikkanen, and Stephen White.

Of those who gave me comments only on Part One, I was helped most by Melissa Barry, David Copp, Joshua Gert, Pamela Hieronymi, Julia Markovits, Sven Nyholm, Connie Rosati, Jeffrey Sebo, David Sobel, Sigrun Svavarsdottir, David Velleman, and Michael Zimmerman.

Of those who gave me comments on my claims about Kant, I was helped most by Marcia Baron, David Cummiskey, Richard Dean, Jeffrey Edwards, Paul Guyer, Thomas Hill, Samuel Kerstein, Patricia Kitcher, Thomas Pogge, and Allen Wood. I have failed to respond adequately to the comments of Edwards, Kitcher, and Pogge on my interpretations of Kant, and to the comments by Samuel Freeman and Leif Wenar on my claims about Rawls.

Of those who gave me comments only on Part Six, I was helped most by Robert Adams, Paul Boghossian, Laurence Bonjour, Nicholas Bostrom, Philip Bricker, Justin Clarke-Doane, Terence Cuneo, Cian Dorr, Kit Fine, Stephen Finlay, Alvin Goldman, Bob Hale, Michael

Jubien, Thomas Kelly, Brian Leiter, William Lycan, Tim Maudlin, Brian McLaughlin, Charles Parsons, Simon Rippon, Stephen Schiffer, Mark Schroeder, Russ Shafer-Landau, Peter Singer, Knut Skarsaune, Robert Stalnaker, Brian Weatherson, Ralph Wedgwood, and Timothy Williamson.

I have also been helped by Larry Alexander, Henry Allison, Gustaf Arrhenius, Elizabeth Ashford, Bruce Aune, Annette Baier, Matthew Bedke, Akeel Bilgrami, Daniel Boisvert, Matthew Boyle, Sarah Buss, Krister Bykvist, Thomas Carson, Timothy Chappell, Daniel Cohen, Joshua Cohen, Robert Curtis, Gordon Davis, Paul Dinkin, Thomas Donaldson, Dale Dorsey, Jamie Dreier, Julia Driver, Jerry Dworkin, Andrew Egan, Nir Eyal, Geoffrey Ferrari, Claire Finkelstein, Katrin Flikschuh, Johann Frick, Jerry Gaus, Berys Gaut, Tamar Gendler, Pablo Gilabert, Margaret Gilbert, George Giovanni, Joshua Glasgow, James Grant, Liron Greenstein, Alex Gregory, Ish Haji, Jason Hanna, Robert Hanna, Joshua Harlan, Daniel Hausman, Allan Hazen, Christopher Heathwood, Dieter Henrich, David Heyd, Alison Hills, Nathan Holcomb, Mike Huemer, Thomas Hurka, Paul Hurley, Susan Hurley, Frank Jackson, Dale Jamieson, Justin Jeffrey, Leonard Kahn, Robert Kane, Stephen Kearns, Paul Klumpe, Richard Kraut, Rahul Kumar, Joel Kupperman, Arto Laitinen, Robin Lawlor, Mark LeBar, James Lenman, John Leslie, Hallvard Lillehammer, Don Loeb, David Lyons, Tienmu Ma, Jacqueline Marina, David McCarthy, Kris McDaniel, Dennis McKerlie, Chris McMahon, David McNaughton, Elijah Millgram, Adrian Moore, Sophia Moreau, Adam Morton, Istvan Musza, Jan Narveson, Stephen Nathanson, William Nelson, Michael Neumann, Kenneth O'Day, Avner Offer, Onora O'Neill, Serena Olsaretti, Jonas Olson, Toby Ord, Leah Orent, Francesco Orsi, David Owens, Stephen Palmquist, Herlinde Pauer-Studer, David Phillips, Christian Piller, Richard Price, Bogdan Rabanca, Wlodek Rabinowicz, Toni Rønnow-Rasmussen, Joseph Raz, Andrews Reath, Bernard Reginster, Michael Ridge, Arthur Ripstein, Michael Rohlf, Gideon Rosen, Mike Rosen, Carol Rovane, Angelica Rudenstine, Julian Savulescu, Jerome Schneewind, Dieter Schoenecker, Frederick Schueler, Bart Schultz, Sally Sedgwick, Jeffrey Seidman, Matthew Seligman, Julius Sensat, Andrew Sepielli, Robert Shaver, Walter Sinnott-Armstrong, John Skorupski, Holly Smith, Michael Smith, Tom

Sorell, Carlos Soto, Amia Srinivasan, Cynthia Stark, Philip Stratton-Lake, Galen Strawson, Bart Streumer, Roger Sullivan, Adam Swenson, Folke Tersman, Jens Timmerman, Torbjörn Tännsjö, Pekka Vayrynen, Edna Ullmann-Margalit, David Velleman, Benjamin Vilhauer, Gerard Vong, Alex Voorhoeve, R. Jay Wallace, James Walmsley, Paul Weirich, Kenneth Westphal, Evan Williams, Chris Woodard, Helena Wright, and Masahiro Yamada.

I thank All Souls College for the immense privilege of a Research Fellowship during the many years in which I have written this book. I thank the Tanner Foundation for supporting the lectures which I expanded into Parts Two and Three. I am grateful to the Commentators on these lectures who wrote Part Four, and to Samuel Scheffler for his work as Editor. I am grateful to Jenny Lunsford, Eleanor Collins, and Clare Hofmann, of the Oxford University Press, for responding generously to my many interferences with the production of these volumes. And I am very grateful to Peter Momtchiloff, my publishing editor, for giving me, over many years, so much wise advice.

SUMMARY

PART ONE REASONS

CHAPTER 1 NORMATIVE CONCEPTS

1 *Normative Reasons*

We are the animals that can both understand and respond to reasons. Facts give us reasons when they count in favour of our having some belief or desire, or acting in some way. When our reasons to do something are stronger than our reasons to do anything else, this act is what we have *most reason* to do, and may be what we *should*, *ought to*, or *must* do. Though it is facts that give us reasons, what we can *rationally* want or do depends instead on our beliefs.

2 *Reason-Involving Goodness*

Things can be good or bad by having features that might give us reasons to respond to these things in certain ways. Events can be good or bad *for* particular people, or *impersonally* good or bad, in reason-implying senses. On some widely accepted views about reasons, nothing could be in these ways good or bad.

CHAPTER 2 OBJECTIVE THEORIES

3 *Two Kinds of Theory*

According to *subjective* theories, we have most reason to do whatever would best fulfil or achieve our present desires or aims. Some Subjectivists appeal to our actual present desires or aims; others appeal to the

desires or aims that we would now have, or to the choices that we would now make, if we had carefully considered the relevant facts. Since these are all facts about *us*, we can call such reasons *subject-given*. According to *objective* theories, we have reasons to act in some way only when, and because, what we are doing or trying to achieve is in some way good, or worth achieving. Since these are facts about the *objects* of these desires or aims, we can call such reasons *object-given*. They are also *value-based*. Theories of these two kinds often deeply disagree. We ought, I shall argue, to accept some value-based objective theory.

4 *Responding to Reasons*

When we are aware of facts that give us strong reasons to have particular desires, our response to these reasons is seldom voluntary. Nor can we choose how we respond to most of our reasons to have particular beliefs. Our rationality consists in part in our non-voluntary responses to these reasons.

5 *State-Given Reasons*

When it would be good if we had certain beliefs or desires, that may seem to give us reasons to have these beliefs or desires. But such reasons would have no importance.

6 *Hedonic Reasons*

The same facts give us object-given reasons both to have and to try to fulfil certain desires. What we want is always some possible event, in the wide sense that covers acts and states of affairs. We have *telic* reasons to want some events as ends, or for their own sake, and *instrumental* reasons to want some events as a means to some good end. We have most reason to do whatever would best achieve the ends that we have most reason to want, because the intrinsic features of these ends make them relevantly best.

When we are in pain, what is bad is not our sensation but our conscious state of having a sensation that we dislike. It is similarly good to have sensations that we like. Such *hedonic likings* or *disliking* cannot be rational or irrational, since we have no reasons to like or dislike these

sensations. We also have *meta-hedonic desires* about our own and other people's pleasures and pains. Such desires or preferences *can* be rational or irrational, since we can have strong reasons to have them. It is our hedonic likings and dislikings, not our meta-hedonic desires, that make these conscious states good or bad; so the examples of pleasure and pain do not support the view that our desires can give us reasons, and can make their objects good.

7 *Irrational Preferences*

If we want some event as an end, but this event's intrinsic features give us strongly decisive reasons to want this event *not* to occur, our wanting this event is contrary to reason, and irrational. It would be irrational, for example, to prefer to have one hour of agony tomorrow rather than one minute of slight pain later today. These claims may seem too obvious to be worth making. But such claims are denied by some great philosophers, and they cannot be made by those who accept subjective theories about reasons.

CHAPTER 3 SUBJECTIVE THEORIES

8 *Subjectivism about Reasons*

Subjectivism takes several forms. Subjective theories may appeal to all of our present telic desires, or only to desires that rest on true beliefs, or only to fully informed desires. Some Subjectivists appeal to the choices that we would now make after informed and rational deliberation. Some Objectivists appeal to the choices that we would make, after such deliberation, if we were rational. Though these claims seem similar, they are very different. These Subjectivists claim only that we should deliberate in ways that are *procedurally* rational. Objectivists make claims about what we would choose if we were *substantively* rational. According to Objectivists, what we *ought rationally* to choose depends on our reasons. According to these Subjectivists, our reasons depend on what, after such deliberation, we *would in fact* choose.

9 *Why People Accept Subjective Theories*

Since so many people believe that *all* practical reasons are desire-based, aim-based, or choice-based, how could it be true that, as objective

theories claim, there are *no* such reasons? How could all these people be so mistaken? There are several possible explanations, since there are several ways in which our desires or aims may seem to give us reasons.

10 *Analytical Subjectivism*

Some claims seem to be *substantive*, but are merely *concealed tautologies*, which everyone could accept whatever else they believe. Several Subjectivists use the words 'reason', 'should', and 'ought' in *subjectivist* senses. These people's theories do not make substantive claims.

11 *The Agony Argument*

Substantive subjective theories can have implausible implications. These theories imply, for example, that we often have no reason to want to avoid some future period of agony. Some Subjectivists would respond to this objection by appealing to claims about procedural rationality. Such replies fail.

CHAPTER 4 FURTHER ARGUMENTS

12 *The All or None Argument*

Subjective theories could also imply that we have decisive reasons to cause ourselves to be in agony for its own sake, to waste our lives, and to try to achieve other bad or worthless aims. In response to this objection, Subjectivists might claim that, for some desire or aim to give us a reason, we must have some reason to have this desire or aim. But these people cannot defensibly make this claim. On subjective theories, all that matters is *whether* some act would fulfil our present fully informed desires or aims. It is irrelevant *what* we want, or are trying to achieve. Either *all* of these desires give us reasons for acting, or *none* of them do. Since it is clear that some of these desires could not give us reasons, we should conclude that none of them do.

Some of our desires can be claimed to give us reasons to have other desires, but any such chain of desire-based reasons must begin with some desire that we have no reason to have. Since such desires cannot be defensibly claimed to give us reasons, Subjectivists cannot defensibly

claim that we have desire-based reasons to have any desire or aim, or to act in any way.

13 *The Incoherence Argument*

Many Subjectivists claim that we have most reason to fulfil, not our actual present desires or aims, but the desires or aims that we would now have if we knew the relevant facts. These people also claim that, when we are making important decisions, we ought to try to learn more about the different possible outcomes of our acts, so that we shall come to have better informed desires. Since Subjectivists deny that the intrinsic features of these outcomes give us reasons, they cannot coherently make these claims.

14 *Reasons, Motives, and Well-Being*

If we are Subjectivists, we must deny that events can be good or bad for particular people, or impersonally good or bad, in the reason-implying senses. When some writers claim that some life would be best for someone, they mean that this is the life that, after fully informed and procedurally rational deliberation, this person would in fact choose. On this account, the best life for someone might be a life of unrelieved suffering. That is not a helpful claim. Some other accounts fail in other ways.

15 *Arguments for Subjectivism*

On subjective theories, *nothing matters*. We should reject the arguments for this bleak view.

CHAPTER 5 RATIONALITY

16 *Practical and Epistemic Rationality*

We are rational insofar as we respond well to reasons or apparent reasons. We have some *apparent* reason when we have beliefs about the relevant facts whose truth would give us some reason. Our desires and acts are rational when, if our beliefs were true, we would have sufficient reasons to have these desires, and to act in these ways. Some people add

that, for our desires or acts to be rational, they must depend on rational beliefs. This claim is misleading, and not worth making.

On one view, what is distinctive of epistemic rationality is the aim of reaching true beliefs. There is another, better view. As well as drawing a deeper distinction between epistemic and practical rationality, we should draw this distinction in a different way, and in a different place.

17 *Beliefs about Reasons*

According to some writers, to be fully rational, we don't need to respond to reasons, or apparent reasons. It is enough to avoid certain kinds of inconsistency, such as failing to respond to what we ourselves believe to be reasons. Such views are too narrow.

18 *Other Views about Rationality*

The rationality of our desires does not depend, as many people claim, on whether these desires are consistent, or on how we came to have them, or on whether our having them has good effects. Our desires are rational when they depend on beliefs whose truth would make the objects of these desires, or what we want, in some way good or worth achieving.

CHAPTER 6 MORALITY

19 *Sidgwick's Dualism*

We can assess the strength of our reasons, Sidgwick seems to argue, from two points of view. When assessed from our personal point of view, self-interested reasons are supreme. When assessed from an impartial point of view, impartial reasons are supreme. To compare the strength of these two kinds of reason, we would need some third, neutral point of view. Since there is no such point of view, self-interested and impartial reasons are *wholly incomparable*. When reasons of these two kinds conflict, neither could be stronger. We would always have sufficient or undefeated reasons to do either what would be impartially best or what would be best for ourselves.

We should reject Sidgwick's argument. We ought to assess the strength of all our reasons from our actual, personal point of view, and we do not need a neutral point of view. We should also revise Sidgwick's conclusion. We have personal and partial reasons to be specially concerned, not only about our own well-being, but also about the well-being of certain other people, such as our close relatives and those we love. These are the people to whom we have *close ties*. We also have impartial reasons to care about anyone's well-being, whatever that person's relation to us. Though there are truths about the relative strengths of these two kinds of reason, Sidgwick's view is partly right, since these comparisons are, even in principle, very imprecise. As *wide value-based objective* theories claim, when one of two possible acts would be impartially better, but the other act would be better either for ourselves or for those to whom we have close ties, we often have sufficient reasons to act in either way.

20 *The Profoundest Problem*

As well as asking 'What do I have most reason to do?', we can ask 'What ought I morally to do?' If these questions often had conflicting answers, because we often had most reason to act wrongly, morality would be undermined. Like other normative requirements, moral requirements matter only when they give us reasons.

Though reasons are more fundamental, much of what follows is about morality. But I shall also be discussing reasons. Several moral principles and theories appeal to claims about what, in actual or imagined situations, we would have most reason or sufficient reason to consent to, or agree to, or to want, or choose, or do.

CHAPTER 7 MORAL CONCEPTS

21 *Acting in Ignorance or with False Beliefs*

By distinguishing several senses of 'ought morally' and 'wrong', we can recognize some important truths and avoid some unnecessary disagreements. Acts can be wrong in *fact-relative*, *evidence-relative*,

belief-relative, and *moral-belief-relative* senses. Facts about these kinds of wrongness provide answers to different questions. When what we ought to do depends on the goodness of our act's effects, we ought to try to do, not what would in fact be best, but what would be *expectably-best*.

22 *Other Kinds of Wrongness*

There are several other senses of 'wrong', which may refer to different kinds of wrongness. Most of these senses are worth using.

It is a difficult question whether, as I believe, there are some irreducibly normative truths, some of which are moral truths. These questions will be easier to answer when we have made more progress in our thinking about practical and epistemic reasons, and about morality. Rather than proposing a new moral theory, I shall try to develop existing theories of three kinds: Kantian, Contractualist, and Consequentialist.

PART TWO PRINCIPLES

CHAPTER 8 POSSIBLE CONSENT

23 *Coercion and Deception*

We act wrongly, Kant claims, when we treat people in any way to which they cannot possibly consent. This claim may seem to imply that we ought never to coerce or deceive people, since these may seem to be acts whose nature makes consent impossible. But that is not relevantly true.

24 *The Consent Principle*

Kant's claims about consent can be interpreted in two ways. On the *Choice-Giving Principle*, it is wrong to treat people in any way to which these people *cannot actually* give or refuse consent, because we have failed to give these people the power to choose how we treat them. This principle is clearly false. On the *Consent Principle*, it is wrong to treat people in any way to which they *could not rationally* consent, if we gave them the power to choose how we treat them. This principle is more likely to be what Kant means, and might be true.

Kant's claims gives us an inspiring ideal of how, as rational beings, we ought to be related to each other. We might be able to treat everyone only in ways to which they could rationally consent; and this might be how everyone ought always to act.

25 *Reasons to Give Consent*

Whether we could achieve Kant's ideal depends on which are the acts to which people could rationally give informed consent, because they would have sufficient reasons to consent. If the best theory about reasons were either some subjective theory, or Rational Egoism, the Consent Principle would fail, since there would be countless permissible or morally required acts to which some people could not rationally consent. But if the best theory is some wide value-based objective theory, as I believe, the Consent Principle may succeed. As some examples suggest, there may always be at least one possible act to which everyone could rationally consent. And we have reasons to believe that, in all such cases, it would be wrong to act in any way to which anyone could not rationally consent.

26 *A Superfluous Principle?*

According to some writers, even if the Consent Principle is true, this principle adds nothing to our moral thinking. What is morally important is not the fact that people could not rationally consent to certain acts, but the various facts that give these people decisive reasons to refuse consent. When applied to acts that affect only one person, this objection has some force. But when our acts would affect many people, if there is only one possible act to which everyone could rationally consent, this fact would give us a strong reason to act in this way, and would help to explain why the other possible acts would be wrong. It is also worth asking whether we could achieve Kant's ideal.

27 *Actual Consent*

It is wrong to treat people in certain ways if these people either do not, or would not, actually consent to these acts. Such acts are wrong even if these people could have rationally given their consent. That is no

objection to the Consent Principle, which claims to describe only one of the facts that can make acts wrong.

On one view, it is wrong to treat people in any way to which they actually refuse consent. That is clearly false. It may seem that no one could rationally consent to being treated in any way to which they actually refuse consent. If that were true, the Consent Principle would also be clearly false. But this objection can be answered.

According to *the Rights Principle*, everyone has rights not to be treated in certain ways without their actual consent. In stating and applying this principle, we would need to answer some difficult questions.

28 *Deontic Beliefs*

To explain why the Consent Principle does not mistakenly require certain wrong acts, we must appeal to the fact that these acts are wrong in other ways, or for other reasons. On some plausible assumptions, the Consent Principle could never require us to act wrongly, because any act's wrongness would give everyone sufficient reason to consent to our failing to act in this way.

29 *Extreme Demands*

The Consent Principle can require us to bear great burdens, when that would save some other people from much greater burdens. If this requirement is too demanding, we would have to revise this principle. But we might still be able to achieve Kant's ideal.

CHAPTER 9 MERELY AS A MEANS

30 *The Mere Means Principle*

It is wrong, Kant claims, to treat any rational being merely as a means. We treat people in this way when we both use these people and regard them as mere tools, whom we would treat in whatever way would best achieve our aims. On a better version of Kant's principle, it is wrong to treat people merely as a means, or to *come close* to doing that.

We do not treat someone merely as a means, nor are we close to doing that, if either (1) our treatment of this person is governed in sufficiently important ways by some relevant moral belief or concern, or (2) we do or would relevantly choose to bear some great burden for this person's sake.

Suppose that some Egoist benefits himself by keeping some promise to someone whose help he needs, and saving some drowning child for the sake of getting some reward. Since this man treats these other people merely as a means, Kant's principle mistakenly condemns these acts. We could qualify this principle, so that it condemns treating someone merely as a means only if our act is also likely to harm this person.

Suppose next that some driverless runaway train is headed for a tunnel in which it would kill five people. These people's lives cannot be saved except by your causing me, without my consent, to fall onto the track, thereby killing me but stopping the train. It may seem that, if you acted in this way, you would be treating me merely as a means. But in some versions of this case that would not be true. And I could rationally consent to being treated in this way. Though such acts may be wrong, that wrongness is not implied by either the Mere Means Principle or the Consent Principle.

31 As *a Means and* Merely *as a Means*

It is widely believed that if we harm people, without their consent, as a means of achieving some aim, we thereby treat these people merely as a means, in a way that makes our act wrong. This view involves three mistakes. When we *harm* people as a means, we may not be treating *these people* as a means. Even if we *are* treating these people as a means, we may not be treating them *merely* as a means. And even if we *are* treating them merely as a means, we may not be acting wrongly.

Some people give other accounts of what is involved in treating people merely as a means. These accounts seem to be either mistaken, or unhelpful.

32 *Harming as a Means*

If it would be wrong to impose certain harms on people as a means of achieving certain aims, these acts would be wrong even if we were *not* treating these people *merely* as a means. And if it would *not* be wrong to impose certain other harms on people as a means of achieving certain aims, these acts would not be wrong even if we *were* treating these people merely as a means. Though it is wrong to *regard* anyone merely as a means, the wrongness of our *acts* never or hardly ever depends on whether we are treating people merely as a means.

CHAPTER 10 RESPECT AND VALUE

33 *Respect for Persons*

We ought to respect everyone, but that does not tell us how we ought to act. It is wrong, some writers claim, to treat people in ways that are incompatible with respect for them. This claim does not help us to decide, in difficult cases, whether some act would be wrong.

34 *Two Kinds of Value*

Some things have a kind of value that is to be *promoted*. Possible acts and other events are in this way good when there are facts about them that give us reasons to make them actual. People have a kind of value that is to be *respected*. Such value is not a kind of goodness.

35 *Kantian Dignity*

Kant uses 'dignity' to mean supreme value or worth. It is sometimes claimed that, on Kant's view, such supreme value is had only by rational beings, or persons, and is the kind of value that should be respected rather than promoted. But that is not Kant's view. There are several ends or outcomes that Kant claims to have supreme value, and to be ends that everyone ought to try to promote.

Some of Kant's remarks suggest that non-moral rationality has supreme value. But Kant's main claims do not commit him to this implausible view. Kant fails to distinguish between being supremely good and having

a kind of moral status that is compatible with being very bad. But we can add this distinction to Kant's view.

36 *The Right and the Good*

Some ancient Greeks, Kant claims, mistakenly tried to derive the moral law from their beliefs about the Greatest Good. But Kant describes an ideal world, which he calls the *Highest* or *Greatest Good*, and he claims that everyone ought always to strive to produce this world. Kant may seem here to be making what he calls the 'fundamental error' of these ancient Greeks. But that is not so.

37 *Promoting the Good*

In Kant's ideal world, everyone would be virtuous and would have all the happiness that their virtue would make them deserve. We can do most to produce this world, Kant claims, by strictly following his other principles. It is often thought that, when Kant claims that lying is always wrong, he is thereby rejecting Act Consequentialism. That is not so. But when Kant, Hume, and others make such claims, they fail to draw some distinctions that we need to draw.

CHAPTER 11 FREE WILL AND DESERT

38 *The Freedom that Morality Requires*

If our acts were merely events in time, Kant argues, these acts would be causally determined, so we could never have acted differently, and morality would be an illusion. Since morality is not an illusion, our acts are not merely events in time. This argument fails. Though we *ought* to have acted differently only if we *could* have done so, the relevant sense of 'could' is compatible with determinism.

39 *Why We Cannot Deserve to Suffer*

According to another of Kant's arguments, if our acts were merely events in time, we could never be responsible for these acts in some way that could make us deserve to suffer. Since we *can* be responsible for our acts in this desert-involving way, our acts are not merely such events.

Though this argument is valid, it is not sound. We ought to accept Kant's claim that, if our acts were merely such events, we could not deserve to suffer. But since we ought to reject this argument's conclusion, we ought to reject Kant's other premise. Our acts *are* merely events in time. So we cannot deserve to suffer.

PART THREE THEORIES

CHAPTER 12 UNIVERSAL LAWS

40 *The Impossibility Formula*

By our *maxims* Kant means, roughly, our policies and underlying aims. According to Kant's *stated* version of what we can call his *Impossibility Formula*, it is wrong to act on any maxim that could not be a universal law. There is no useful sense in which this could be claimed to be true.

According to Kant's *actual* version of his Impossibility Formula, it is wrong to act on any maxim of which it is true that, if everyone accepted and acted on this maxim, or everyone believed that they were morally permitted to act upon it, that would make it impossible for anyone successfully to act upon it. This formula spectacularly fails, since it does not condemn acts of self-interested killing, injuring, coercing, lying, and stealing. Kant's formula rightly condemns the making of lying promises. But this formula condemns such acts for a bad reason, and it mistakenly condemns some good or morally required acts.

41 *The Law of Nature and Moral Belief Formulas*

Kant proposes another, better formula. To apply this formula, we suppose that we have the power to *will*, or choose, that certain things be true. We act wrongly, Kant claims, if we act on some maxim that we could not rationally will to be a universal law. There are three versions of this *Formula of Universal Law*. According to

> *the Law of Nature Formula*, it is wrong to act on some maxim unless we could rationally will it to be true that everyone accepts this maxim, and acts upon it when they can.

According to

> *the Permissibility Formula*, it is wrong to act on some maxim unless we could rationally will it to be true that everyone is morally permitted to act upon it.

According to

> *the Moral Belief Formula*, it is wrong to act on some maxim unless we could rationally will it to be true that everyone believes that such acts are morally permitted.

It will be enough to consider Kant's Law of Nature and Moral Belief Formulas. These formulas develop the ideas that are expressed in two familiar questions: 'What if everyone did that?' and 'What if everyone thought like you?'

When we apply these formulas, we must appeal to some view about rationality and reasons. Since we are asking what Kant's formulas can achieve, we should appeal to what we believe to be the best view. But we should not appeal to our beliefs about which acts are wrong, or to the *deontic* reasons that such wrongness might provide, since Kant's formulas would then achieve nothing.

42 *The Agent's Maxim*

Whether some act is wrong, Kant's formulas assume, depends on the agent's maxim. Most of the maxims that Kant discusses are, or include, *policies*. Suppose that some Egoist has only one maxim or policy: 'Do whatever would be best for me'. This man could not rationally will it to be true either that everyone acts on this maxim, or that everyone believes such acts to be permitted. Most Egoists could not rationally choose to live in a world of Egoists, since that would be much worse for them than worlds in which people act on various moral maxims. Whenever our imagined Egoist acts on his maxim, Kant's formulas imply that this man's acts are wrong. This man acts wrongly even when, for self-interested reasons, he pays his debts, puts on warmer clothing, and saves some drowning child in the hope of getting some reward. These implications are clearly

false. When this Egoist acts in these ways, his acts do not have what Kant calls *moral worth*, but they are not wrong.

Consider next Kant's maxim 'Never lie'. Kant could not have rationally willed it to be true that no one ever tells a lie, not even to a would-be murderer who asks where his intended victim is. Kant's formula therefore implies that, if Kant acted on this maxim by telling anyone the truth, he acted wrongly. That is clearly false. As these and other cases show, whether some act is wrong cannot depend on the agent's maxim, in the sense that can refer to policies. There are many policies on which it is sometimes but not always wrong to act. Nor does an act's moral worth depend on the agent's maxim.

Kant's appeal to the agent's maxim raises other problems. Such problems have led some people to believe that Kant's Formula of Universal Law cannot help us to decide which acts are wrong. When used as such a criterion, these people claim, Kant's Formula is unacceptable, worthless, and cannot be made to work.

Kant's Formula *can* be made to work. When revised in certain ways, I shall argue, this formula is remarkably successful.

Some writers suggest that, rather than appealing to the agent's actual maxim, Kant's Formula should appeal to the possible maxims on which the agent might have been acting. This suggestion fails.

In revising our two versions of Kant's Formula, we should drop the concept of a maxim, and use instead the morally relevant description of the acts that we are considering. The Law of Nature Formula could become:

> We act wrongly unless we are doing something that we could rationally will everyone to do, in similar circumstances, if they can.

The Moral Belief Formula could become:

> We act wrongly unless we could rationally will it to be true that everyone believes such acts to be permitted.

These formulas will need some further revisions.

It may be objected that, if we revise Kant's formulas by dropping the concept of a maxim, we are no longer discussing Kant's view. That is true, but no objection. We are developing a Kantian moral theory, in a way that may make progress.

CHAPTER 13 WHAT IF EVERYONE DID THAT?

43 *Each-We Dilemmas*

It will be simpler to go on discussing Kant's formulas, returning to our revised versions when that is needed.

On Kant's Law of Nature Formula, it is wrong to act on some maxim unless we could rationally will it to be true that *everyone* rather than *no one* acts upon it. We are often members of some group of whom it is true that, if *each* rather than *none* of us did what would be *better* for ourselves, *we together* would be doing what would be *worse* for all of us. In many such cases, each of us could either benefit ourselves or give some greater benefit to others. We can face similar *each-we dilemmas* when we have certain other morally permitted or required aims, such as the aim of promoting our children's well-being. It may be true that, if each rather than none of us did what would be better for our own children, *we* would be doing what would be worse for everyone's children. We could not rationally will it to be true that everyone rather than no one acts in these ways. So if everyone followed Kant's Law of Nature Formula, no one would act in these ways, and that would be better for everyone. These are the cases in which we can best think and say 'What if everyone did that?'

Kant's formula is especially valuable when the bad effects of any single act are spread over so many people that the effects on each person are trivial or imperceptible. One example are the acts with which we are selfishly over-heating the Earth's atmosphere. By requiring us to do only what we could rationally will everyone to do, Kant's formula helps us to see how much harm we are doing, and strongly supports the view that such acts are wrong. In some of these cases, we can add, common sense morality is *directly collectively self-defeating*, and should therefore be revised.

44 *The Threshold Objection*

Whether it is wrong to act on some maxim sometimes depends on how many people act upon it. There are some maxims on which it is permissible or good for some people to act, though it would be very bad if everyone acted on them. Two examples are the maxims 'Consume food without producing any,' and 'Have no children, so as to devote my life to philosophy'. Most of us could not rationally will it to be true that everyone acts on these maxims, so Kant's Law of Nature Formula condemns such acts even when they are not wrong. This objection is partly answered by the fact that most people's maxims implicitly take into account what other people are doing. For a complete answer, we must revise Kant's formula.

45 *The Ideal World Objections*

Kant's Law of Nature Formula, it is often claimed, requires us to act as if we were living in an ideal world, even when in the real world such acts would have predictably disastrous effects and be clearly wrong. We are required, for example, never to use violence even in self-defence, and required to act in various ways that mistakenly ignore what other people will in fact do. This *Ideal World Objection* can be answered. Kant's formula does not require such acts.

There is a different problem. Once a few people have failed to do what we could rationally will everyone to do, Kant's formula permits the rest of us to do whatever we like. Similar objections apply to some *Rule Consequentialist* moral theories. To answer this *New Ideal World Objection*, we should revise Kant's formula in another way. It is wrong to act on some maxim, this formula could claim, unless we could rationally will it to be true that this maxim be acted on, not only by everyone rather than by no one, but also by *any other number* of people rather than by no one. Rule Consequentialists could make similar claims.

Of the two versions of Kant's Formula of Universal Law, the Moral Belief Formula is better. When people object 'What if everyone did that?', it is often enough to reply 'Most people won't'. But when people object 'What if everyone thought like you?', it is *not* enough merely to reply 'Most people won't'.