READING
PARFIT
ON WHAT MATTERS

EDITED BY
SIMON KIRCHIN

READING PARFIT

On What Matters

Edited by Simon Kirchin

Routledge
Taylor & Francis Group

LONDON AND NEW YORK

CONTENTS

CONTENTS

CONTRIBUTORS

David Copp is Distinguished Professor of Philosophy at the University of California, Davis. He is the author of *Morality, Normativity, and Society* (Oxford University Press, 1995) and *Morality in a Natural World* (Cambridge University Press, 2007) and he has edited several anthologies, including *The Oxford Handbook of Ethical Theory* (Oxford University Press, 2006). He is the editor of a monograph series with Oxford University Press called *Oxford Moral Theory*. He has published and lectured widely on topics in moral and political philosophy.

J.L. Dowell, an Associate Professor at Syracuse, has published in metaphysics, philosophy of language, philosophy of mind, philosophical methodology and metaethics. She is the 2014 winner of the Marc Sanders Prize in Metaethics and is currently working on a book on deontic modals, under contract with Oxford University Press.

Julia Driver is Professor of Philosophy at Washington University in St. Louis. She works primarily in normative ethics and moral psychology. She has published three books: *Consequentialism* (Routledge, 2012), *Ethics: the Fundamentals* (Blackwell, 2006), and *Uneasy Virtue* (Cambridge University Press, 2001).

Simon Kirchin is Reader in Philosophy at the University of Kent and is currently Dean of its Faculty of Humanities. He works mainly in ethics and metaethics and is a past President of the British Society for Ethical Theory. He is the author of *Metaethics* (Palgrave, 2012) and has edited the following volumes: *Thick Concepts* (Oxford University Press, 2013), *A World without Values* (with Richard Joyce; Springer, 2010), and *Arguing about Metaethics* (with Andrew Fisher; Routledge, 2006). He is currently finishing a monograph with the working title *Thick Evaluation*.

David McNaughton is Professor of Philosophy at Florida State University and Professor Emeritus at Keele University. He is the author of *Moral Vision* (Blackwell, 1988), *Forgiveness* (with Eve Garrard; Routledge, 2010) and of a number of papers on ethics, philosophy of religion and the relations between

the two. He has recently finished editing Joseph Butler's moral writings for Oxford University Press and is currently writing a book with Piers Rawling on their approach to practical reasons.

Julia Markovits is Associate Professor of Philosophy at Cornell University. She joined the Cornell faculty in 2014, after spending five years on the philosophy faculty at MIT. Before that, she was a Junior Fellow at the Harvard Society of Fellows. Markovits researches and teaches in meta- and normative ethics, bioethics and the philosophy of law. More specifically, she has written about questions concerning the nature of moral reasons and about moral praiseworthiness and blameworthiness. She is the author of *Moral Reason* (Oxford University Press, 2014).

Derek Parfit (1942–2017) was for many years a Fellow of All Souls College, Oxford, retiring as Senior Research Fellow in 2010, whereupon he became an Emeritus Fellow. He also held visiting professorships at New York University, Harvard and Rutgers.

Douglas W. Portmore is Professor of Philosophy in the School of Historical, Philosophical and Religious Studies at Arizona State University. He works mainly in moral philosophy. He is the author of *Commonsense Consequentialism: Wherein Morality Meets Rationality* (Oxford University Press, 2011). He is currently working on another book with the working title *Opting for the Best: Oughts and Options*. The book concerns what is, perhaps, the least controversial normative principle concerning action: you ought to perform your best option – that is, the option that is most highly favoured by your reasons.

Piers Rawling is Professor and Chair of Philosophy at Florida State University. He has wide-ranging interests and has published papers on decision theory, ethics (with David McNaughton), philosophy of language, various other areas of philosophy, and quantum computing (with Stephen Selesnick). He is co-editor (with Alfred Mele) of *The Oxford Handbook of Rationality* (Oxford University Press, 2004) and is currently writing a book with David McNaughton on their approach to practical reasons.

Kieran Setiya is Professor of Philosophy at MIT. He works in action theory, ethics and epistemology and is the author of *Reasons without Rationalism* (Princeton, 2007) and *Knowing Right from Wrong* (Oxford University Press, 2012). His essays range from the nature of rational agency to the place of love in moral philosophy and the resolution of the midlife crisis.

David Sobel is Guttag Professor of Ethics and Political Philosophy at Syracuse University. His *From Valuing to Value*, a collection of his old and new work on subjective accounts of well-being and reasons for action, is forthcoming from Oxford University Press. He co-edits, with Peter Vallentyne and Steve Wall, *Oxford Studies in Political Philosophy*.

Philip Stratton-Lake is Professor of Philosophy at Reading University. He is the author of *Kant, Duty, and Moral Worth* (Routledge, 2000) and editor of *Ethical Intuitionism: Re-evaluations* (Clarendon, 2002), W. D. Ross's *The Right and the Good* (Clarendon, 2002) and *On What We Owe to Each Other* (Blackwell, 2004).

ACKNOWLEDGEMENTS

INTRODUCTION

Simon Kirchin

Derek Parfit's *On What Matters* is a striking intervention into modern moral philosophy and was, it is fair to say, one of the most eagerly anticipated works of analytic philosophy published for a long time.

Parfit published *Reasons and Persons*, his first book, in 1984. This influenced a whole generation of thinkers, both within moral philosophy and far beyond it, in its arguments, its ideas and its style of working through philosophical problems. As such, whatever book Parfit published next would have found itself in the spotlight. However, *On What Matters* (hereafter OWM) deserves to be considered and admired on its own terms and for its own reasons. There are a number of distinctive and arresting views that Parfit articulates within its covers, with many topics discussed and numerous arguments offered that range from the subtle to the direct. Indeed, it is probably worth lingering on one detail. Although we may talk of OWM as *a* book, it is a book that, when it was originally published, came in two volumes that ran to just over 1,400 pages (a third volume was published in 2017).[1] Further, it is split into six parts (one comprising commentary from Barbara Herman, T. M. Scanlon, Susan Wolf and Allen Wood) plus appendices. One can justly describe it as 'a work' that is, in fact, a few books.

In this short introduction I do no more than offer a flavour of the topics and ideas that Parfit covers in OWM 1 and 2, roughly in the order in which he discusses them, whilst also summarizing the chapters in this volume.

Parfit begins by thinking about reasons. For him a reason is something conceptually fundamental, something that cannot be explained in, or reduced to, further terms and concepts, even if one can get a sense of what a reason is from various examples and by seeing how it sits with similar normative and evaluative terms and concepts. His key aim throughout Part One is to argue against subjective theories of reasons and to argue in favour of objective theories. Subjectivists about reasons think that what we have most reason to do is (solely) a function of our desires and aims. These may be our actual and present desires and aims, or some desires and aims we would have if we more carefully considered the

1

known facts or were made aware of facts that we do not know. In contrast, objectivists about reasons think that what we have most reason to do is (solely) a function of the facts. For example, we may well have most reason to act in a particular way because it is this action that will bring about the most good. It is clear that subjectivism and objectivism will deliver different conceptions of what we have most reason to do and clear how they can diverge in their final recommendations across a number of situations. For instance, whilst you may think you have most reason to choose a certain career path because it is what you desire to do or be, in fact choosing a different career path would produce the most good. In this case, at least as described in this bare manner, subjectivism and objectivism would differ as to what you have most reason to do.

There are a number of arguments that Parfit offers against subjectivism, some of which parallel his thoughts in Part Six. One line of argument begins by simply stating that subjectivists need to ensure that they are making substantive claims about reasons. They can fall into the danger of dealing in concealed tautologies, moving from the target phrase to be understood (1) 'we have most reason to act in some way' to the phrase (2) 'this act would best fulfil our present fully informed telic desires' (and hence giving sense to 'reason') and then giving a spin on this latter phrase by saying that (3) 'we have most reason to do what would best fulfil our present fully informed telic desires'.[2] If subjectivists use 'have most reason' in the desire-fulfilment sense, then (3) is shown to be a concealed tautology, not a substantive claim: 'the act that would best fulfil our present fully informed telic desires is the act that would best fulfil these desires'. So subjectivists need to use words such as 'reason' in a normative sense and not just as a synonym for the descriptive or factual 'what is most desired'. This immediately creates trouble for them. We can construct scenarios involving the adoption of a course of action in which agents suffer a large amount of pain but where, for whatever reason, they desire to suffer in this way. Subjectivists are then committed to saying that there is most reason (in the normative, substantial sense) for the agent to adopt such a course of action, even when it seems obvious that experiencing such pain is dangerous, bizarre or just plain bad. It strains credulity to think that we would really, sensibly want to say that the agent has most reason to choose this course in most scenarios, and thus subjectivism fails.

Taking himself to have established objectivism's truth through a number of arguments, Parfit moves, in Parts Two and Three, to consider how normative ethicists might seek to advise us as to what we should do.[3] What principles and theories should we adopt in deciding what reasons we, in fact, have? His thoughts here are arguably the single most important contribution that OWM makes to modern debate. In the words of Samuel Scheffler, from his introduction to the whole work, "Parfit aims to rechart the territory of moral philosophy".[4]

Students and scholars alike routinely think that the normative ethical theories of consequentialism and Kantian deontology offer fundamentally different views of what we should do in our moral lives. Consequentialists are typically cast as thinking that the rightness of one's actions is (solely) a function of their

consequences. In contrast, Kantian deontologists are typically cast as eschewing consequences and favouring instead a set of principles or maxims that forbid and encourage certain action-types in accordance with the overarching idea or ideas expressed by the Categorical Imperative. So, to take a simple example, we should not lie because lying treats another person as a means to an end.

Across Parts Two and Three Parfit challenges the assumption that we have fundamental opposition and argues instead that normative ethics contains far more unity than most assume. In order to do this he further refines the theories he is interested in, arguing that his refinements present the best of the broad positions that are part of normative ethics.

He deals with three positions: rule consequentialism, Kantian deontology and contractualism, specifically Scanlonian contractualism. He argues that these three positions will recommend and justify the same, more specific moral principles and actions, and blends them into what he calls the *Triple Theory*:

> TT: An act is wrong just when such acts are disallowed by some principle that is optimific, uniquely universally willable, and not reasonably rejectable.

He goes on to say:

> We can call these the *triply supported* principles. If some principle could have any of these three properties without having the others, we would have to ask which of these properties had most importance. But these three properties, as I have argued, are had by all and only the same principles.[5]

To be clear, Parfit does not advocate that by coincidence these three positions pick out all of the same specific moral principles. Rather, there is something about the nature of these theories and the high-level principles and ideas that are at their core which means they converge on the same specific principles. He thinks there are good reasons to believe the Triple Theory to be true.

Parfit focuses in Parts Two and Three on engaging with Kant's philosophy, Kant being one of the philosophical heroes of *OWM*. Despite his admiration for Kant, Parfit reworks Kant's position, often in radical ways. He rejects or reimagines many points that some commentators think of as central to Kantianism, most notably (I think) the notion of a maxim.[6] A maxim is assumed to be, roughly, the subjective principle or policy on which agents act. 'Subjective' here does not mean what it means above: we are not discussing desire-based principles. Nor is 'subjective' synonymous with 'relativistic'. 'Subjective' here means that something is primarily the agent's. Maxims are those fundamental aims and policies that guide the agent's actions or, to use a shorthand, they are the fundamental motives of the agent that help to explain – indeed help to constitute – his or her action. Kant thinks, roughly, that we can judge the wrongness of the act by whether the maxim

can be universalized. However, there are notorious problems with this. First, if one makes the maxim narrow and detailed (Parfit's example is the theft of a wallet from a woman dressed in white who is eating strawberries whilst reading the last page of Spinoza's *Ethics*), then one can easily universalize without fear that anyone else will act in this way, thus providing oneself with an exception. Yet the action is clearly wrong. In contrast, some maxims are 'mixed', often because they are worded more generally: for example, 'Do what is best for me' and 'Never lie'. Sometimes acting on such maxims can be wrong, but often not, and Kant failed to account for this, according to Parfit. Parfit attempts to show through various examples that the best version of Kantian deontology should eschew maxims, at least on one understanding of that term. We should instead focus on what the morally relevant description of the action is. Focusing on what people are intentionally doing in a particular circumstance will help us to get at such a description, suggests Parfit. For example, in the first example above the person is intending to steal; the other details are irrelevant. In a different case, although I am doing what is best for me by putting on a jumper, I am doing so only to keep warm and hurting no one in the process. Acting in this way can hardly be considered to be wrong. And so on.

As mentioned, Part Four sees four thinkers engage with Parts Two and Three. Susan Wolf claims that in arguing for the Triple Theory Parfit misses much that is of value within the various theories he tries to bring together, for their differences are essential and important to them. Allen Wood raises profound worries about Parfit's philosophical methodology and also disagrees with him about Kant. Whereas Parfit thinks that the Formula of Humanity is not a practically useful principle, Wood disagrees. Barbara Herman also focuses on Parfit's Kantian exegesis, with much of her discussion revolving around the relation between an agent's motive and an act's effects on others. Whilst she is not against trying to see connections and even combinations between theories, she thinks Parfit goes too far in ignoring the importance of motives to the moral worth of actions and brings into question how we arrive at a morally relevant description of an action. Lastly, Scanlon claims that he is not a Kantian and that his position cannot be subsumed into the Triple Theory. He concludes that Parfit takes the production of optimific results to be most morally basic, whereas he himself thinks that what is most morally basic is agreement amongst people. Despite his discussions, thinks Scanlon, Parfit does not capture this type of agreement in the right sort of way. Part Five sees Parfit engage with these four colleagues in which he deepens his view, especially with regards to Scanlon's criticism. He argues that his recasting of Scanlon's view provides Scanlon with a more plausible theory that in turn makes possible the Triple Theory as Parfit conceives it.

This brief summary of Parts Four and Five comes nowhere near doing justice to the material therein and the differing viewpoints one finds. Whilst the details undoubtedly matter, it is worth stressing two themes that emerge strongly from these parts. First, the critics worry that Parfit's position is too consequentialist (that is, too concerned with the production of results) to accommodate the

insights of the other theories satisfactorily. Parfit profoundly disagrees with this, arguing that the best versions of the other two theories are more concerned with the production of moral effects than many people acknowledge. Second, Parfit may well think as he does because of his style of moral reasoning: the main concern of moral philosophy, it seems, is to develop principles to guide our specific actions across all situations. We often refine such principles in the light of the results we get in certain situations (real or imagined) that we test them against. Wood in particular doubts whether this is the best way of proceeding.

In Part Six Parfit switches tack away from normative ethics and towards metaethics. He is a realist and cognitivist about value and normativity and also a staunch non-naturalist. So, for him, normative properties exist and can be things we can know. Furthermore, they cannot be reduced to natural phenomena that are, for example, studied by the natural and social sciences. They are *sui generis*. Parfit considers a number of metaethical positions and writers that seek to offer alternative views, and he argues against all of them. The leading three opposing views are all Ns: noncognitivism, naturalistic realism (both analytic and non-analytic) and nihilism (which incorporates error theory).

Like many other philosophical areas, metaethics has a huge amount of detail and complication as well as a number of chief positions that compete against each other to explain roughly the same phenomenon. What is refreshing about Parfit's Part Six, in my view, is that much of the detail and complication is stripped away. He looks at the essential bones of each position in an attempt to make progress.

Parfit begins by echoing his thought from Part One. He argues that we have external reasons for acting – reasons that do not depend for their existence on any agent's desires or aims – and against those who think that the only reasons that exist are internal reasons – reasons that do so depend. He then moves to provide a battery of ideas and arguments against the positions listed above. A notable argument – the Triviality Objection – employed against non-analytic naturalism mirrors one from Part One.[7] To say that we ought to do something is to make a substantive normative claim. Non-analytic naturalism renders such claims trivial. How so? Take the following claim, which appears to be philosophically substantive:

(U) when some act would maximize happiness, this act is what we ought to do.

U can be claimed by all sorts of utilitarians. Non-naturalist utilitarians such as Sidgwick (the other main philosophical hero in OWM) would claim that the property of maximizing happiness makes the act have the different or further property of being what we ought to do. Naturalist utilitarians claim that the maximization of happiness is the same property as the property of being what we ought to do. If this latter identification is made, says Parfit, then it renders a seemingly substantive claim such as U trivial, for we are then saying only that

when some act would maximize happiness it is an act that would maximize happiness.

At the heart of this move is Parfit's general idea that some other metaethicists incorrectly conceive the subject matter they are trying to explain first of all. If one does not start in the right way, then one can be led into all sorts of failure, as Parfit attempts to show throughout Part Six. Fellow thinkers may render seemingly substantive claims trivial, as above. Or they may fail to explain what it is to disagree with others or how we can improve morally. (These are ideas he raises against noncognitivism.) Or they may have a curious account of reason that fails to do justice to our ethical lives and intuitions. (This is Parfit's main worry with Bernard Williams' thought, echoing the ideas of Part One.) Parfit's overriding concern is that unless one adopts the sort of cognitivist non-naturalism he espouses, then one cannot capture the idea that life and our existence matter, and it is surely right that we do this.

Throughout the whole of OWM there is a boldness in style and orientation which receives two main expressions. In Parts One and Six, where Parfit deals with conceptions of reasons and normativity, he presents an uncompromising account of the reality of the moral and the practical, and of what it takes for things to matter. In Parts Two, Three and Five he is similarly bold. In fashioning a position that seeks to remodel three main normative ethical theories so as to bring them together, he stakes out a position that shakes up the theoretical landscape. In doing so, he begins to give us some idea of how we can decide which things matter ethically. Given the boldness of these aims we will undoubtedly have to measure the success of OWM over a long period of time.

What of the commentators in this book?[8] We begin with my chapter. I discuss the commentary of Wolf and Wood, and Parfit's replies to them. I restate and further Wolf's criticism that the Triple Theory overlooks or unjustly eschews much that is valuable in the three theories Parfit considers. In doing so, I consider the few comments that Parfit makes in his defence. I then turn to Wood's attack on Parfit's philosophical methodology, in part because it strikes at the heart of Parfit's project, and also because Parfit himself prefers to focus on Kant in his reply. I bring the themes from both commentators together (whilst acknowledging their differences), showing how they can lend support to the other, developing points that Parfit needs to answer in order to show that the Triple Theory, or anything like it, is plausible.

Next up is David Copp's chapter. Normative naturalists hold that normative properties and facts are natural, contending that these are similar in all metaphysically important respects to other natural properties and facts. Parfit argues, however, that if normative naturalism were correct, normativity would be illusory and that normative naturalism is close to nihilism. Parfit's most direct argument for this is his Soft Naturalist's Dilemma. From this he concludes that normative naturalists are committed to Hard Naturalism. According to Hard Naturalism, we could get rid of normative terms and concepts without any cognitive loss. Copp argues that this is wrong, focusing on the idea that the naturalist can say,

for example, that even if the property *wrongness* is identical to a natural property, N-wrongness, the proposition that torture is wrong is distinct from the proposition that torture is N-wrong. Hence if we lacked the normative concept of wrongness, there are true propositions we would be unable to believe, including the proposition that torture is wrong. Indeed, we would be unable to formulate the thesis of normative naturalism. There would be a cognitive loss. There is, moreover, argues Copp, additional complexity in our 'ways of thinking' (WOTs) of normative properties that is crucial to normative belief playing its characteristic role in motivating action. To explain this, Copp introduces the idea of an internal WOT of a normative property where 'internally represented' normative beliefs have a characteristic bearing on the motivation of action. This is a feature of internally represented normative beliefs that is not had by naturalistic beliefs, not even if the naturalistic belief has the same truth conditions as the internally represented belief. The upshot is that the naturalist can reject Hard Naturalism. Properly understood, Copp concludes, naturalism does not eliminate normativity. It aims to explain what normativity consists in.

In her contribution Julia Markovits notes the consensus-building of Parts Two and Three and contrasts it with Parfit's total rejection of subjectivism in Part One. She argues both that the difference between objectivism and subjectivism may not be as deep as Parfit presents and that any consensus-building should push us towards subjectivism. A crucial part of her project is to argue that we can have reasons for our desires and that identifying these is a collective project. This leads to an 'optimistic subjectivism', whereby we attempt to identify aims and goals we all have reasons to share, where such reasons are based on desires that we have in common.

In Part One Parfit mentions in passing his commitment to a buck-passing account of goodness, although he disagrees with Scanlon, its most notable defender. In short, Parfit endorses the positive thesis of buck-passing (roughly, that if X is good, then the properties that make X good give us various reasons to act in relation to X) but denies the negative thesis (that goodness itself is never reason-providing). In his contribution, Philip Stratton-Lake also considers buck-passing, and focuses in great detail on the refinement Parfit makes. He also discusses work on this topic by Mark Schroeder. The best case to be made on behalf of Parfit is of understanding X's goodness as a non-additive reason. Stratton-Lake argues that Parfit's view fails and that there is as yet no good reason to reject the negative thesis.

David McNaughton and Piers Rawling in their wide-ranging, joint chapter concern themselves with an overarching idea that emerges across all of *OWM*, namely Parfit's 'two-tier' view of practical reasoning. According to this view, practical reasons are cast as facts. Consider, for example, the following: the fact that you are hungry is a reason to eat some food. There are two facts here, hence why it is two-tier: the fact that you are hungry and the fact that you being hungry is a reason. McNaughton and Rawling trace Parfit's thought across a variety of topics: for example, whether normative notions other than reasons can be central

and irreducible, and the issue of moral constraints in normative theory. This leads them to argue that Parfit should not be a constructivist about morality and should adopt a thoroughgoing non-constructivist two-tier theory.

Kieran Setiya focuses on Parfit's Kantian Contractualism – a crucial part of the Triple Theory – and asks how and whether it can guide action. Kantian Contractualism states that 'everyone ought to follow the principles whose universal acceptance everyone could rationally will'. This provides us with a clear sense of which acts are wrong: an act is wrong if it is deemed wrong by those principles that one accepts under this formula. Through a series of moves, most notably a focus on the idea of a Wrong-Making Reason, Setiya worries whether we can apply the Kantian Contractualist formula when we do not already know what we have reason to do. The formula may be redundant.

Doug Portmore makes trouble for rule consequentialism, another key element of Parfit's Triple Theory. He casts rule consequentialism as stating that agents have reason to act so long as the act is part of a set of acts that, if realized, would bring about the best consequences, and that this is so even if (1) the act itself does not have good consequences and (2) the agent cannot see to it that the set of acts (and their consequences) are realized. Portmore argues that an agent has reason to perform the act only if she can see to it that the set of acts and the consequences are also realized, thus denying (2). This then leads, absent any other sufficient reason to act, to the fact that agents lack sufficient reason to act. So, argues Portmore, this means either that rule-consequentialism is false, or that we often lack sufficient reason to act as morality requires. Both of these options damn Parfit's position.

In their joint chapter J.L. Dowell and David Sobel consider Parfit's argument – the Triviality Objection – against non-analytic naturalism (as considered by Copp). They argue that naturalism can meet the central challenge that Parfit offers. Non-analytic naturalism *can* make informative identity statements, and Parfit misses this because he relies on the mistaken assumption that the informativeness of such statements must be explained by their semantics rather than by the pragmatics of their use. Dowell and Sobel show that it is possible for non-analytic naturalists to make informative identity statements, and hence Parfit's objection is undermined.

Having raised a worry with Parfit's anti-naturalist stance and also considered one of his anti-naturalist arguments, we then change tack to consider what a naturalistic alternative might look like. In her contribution Julia Driver argues for a type of naturalism, whilst taking seriously Parfit's view that metaethical theories should ensure that they can make sense of things mattering. Her approach is broadly Humean. Within this broad approach she defends a view of 'constitutivism', which sees reasons as extractable from basic norms of agency. This can, of course, mean that the reasons that exist are contingent on features of humans and our agency, and this contingency may be unpalatable for certain realists, including, one can imagine, Parfit. Driver argues that this contingency does not in any way lead to a vicious arbitrariness and that this position can still make plausible

sense of why it is that things matter. In this way it ties in nicely with Markovits' chapter.

At the end of the book Parfit replies to all our commentators, with the replies having varying lengths. I do not go into detail here about Parfit's replies. Two main points are worth highlighting, however. First, as one might expect, many of the themes from above appear: the nature and ground of reasons, the status and value of the Triple Theory, practical rationality, Parfit's arguments against naturalism, and others. Second, his replies are robust in his defence, although, as one would expect from Parfit's work, he is always at pains to ensure he gives as clear an answer as possible. It is also worth noting that where he agrees with a fellow writer, Parfit sometimes merely records this fact, whilst at other times he spells out why he thinks that a supposed disagreement is nothing of the sort and why there is some deeper agreement between himself and the writer he is responding to.

This searching for agreement has become a theme in Parfit's writing of late; the advance of the Triple Theory itself shows this. The writers in this volume hope that the various criticisms and ideas discussed here will help to show what (seeming) disputants can in fact agree on and help to underline what can remain as real disagreement.

Notes

All references to *On What Matters* in this volume are referenced as either *OWM* 1 or *OWM* 2, with the relevant page number, chapter or section. *Reasons and Persons* is referenced as *RP*.

1   In this collection we deal only with the first two volumes.
2   *OWM* 1, p. 72.
3   These two parts were first delivered, in different form, as Tanner Lectures in 2002.
4   *OWM* 1, p. xix.
5   *OWM* 1, p. 413.
6   See especially *OWM* 1, §42.
7   *OWM* 2, §95.
8   The order of commentators was suggested by Parfit himself because of how he wanted to structure his responses.

1

REFLECTIONS FROM WOLF AND WOOD

Incommensurability, guidance and the 'smoothing over' of ethical life

Simon Kirchin

Introduction

In *On What Matters* (hereafter *OWM*) Derek Parfit argues that the best versions of Kantianism, Scanlonian contractualism and rule consequentialism can be combined into a position – the Triple Theory – that shows us what sort of ethical principles we should adopt to guide our behaviour and moral judgement.

These three theories are traditionally thought to be rivals, with deep differences. The prospect of their convergence is one of Parfit's most exciting proposals in *OWM*. In this chapter I think about the very idea of combining these three theories. I do so by looking at Parfit's ambitions through the eyes of two of his commentators from Volume 2, namely Susan Wolf and Allen Wood.[1]

This may seem an odd step in a volume devoted to Parfit's work. But I do so because, in his interesting responses, Parfit doesn't engage with what I find most arresting about what Wolf and Wood say. Their criticisms connect with the heart of the whole *OWM* project, and part of my aim is to encourage Parfit to say something in his defence.

Wolf suggests that the attempt to synthesize Kantianism, Scanlonian contractualism and rule consequentialism is unwise, mainly because these theories see different features of our lives as being ethically significant and because they cast many of the same moral features differently. Having highlighted particular parts of Wolf's criticism, I extend her commentary by articulating the theoretical underpinning that Parfit seems to assume for his view. He assumes that normative ethical theories are good and decent only if they can provide clear, practical guidance, and in turn this requires an assumption that all values and things valued are commensurable. This has, in addition, connections with his metaethics in Part Six.

I believe his meta-normative ethic – that is, his theory about what normative ethics is about and how it should be conducted – is essential to the advancement of his Triple Theory, and yet it gets little if any articulation in OWM and certainly no detailed defence.

What of Wood? He criticizes Parfit's methodology, amongst other matters. Whilst both he and Parfit are interested in practical guidance, I use the differences between their methodologies and conceptualizations to illustrate and deepen Wolf's concern. Whilst that is a prime aim of mine, I also repeat and extend some of Wood's ideas to, again, encourage Parfit to reply.

Doubt is cast by both commentators not so much on the details of the Triple Theory itself but on Parfit's more general hope of drawing together much of what is important in the Western moral canon in order to advance our moral thought. What is embodied in the Triple Theory may have more narrow appeal and success than Parfit seems to think.

Wolf

In summarizing and discussing Wolf's ideas in this section and the next, I emphasize and extend three interrelated themes: incommensurability, the conception of action guidance offered by normative ethics and how these first two ideas relate to Parfit's concerns about disagreement and reality.

Wolf begins her rich and interesting commentary by articulating Parfit's ambition in OWM. It is not just that Parfit is picking and choosing what he takes to be best in the three main theories he focuses on. He aims, too, to systematize them individually and then synthesize them to show us that, perhaps imperfectly, proponents of these views are attempting to reach the same single true morality. Parfit shares the assumption or hope that there is a single true morality with "many if not all of the major figures in the traditions he claims to combine".[2] For Wolf, in contrast, it would not be such a "moral tragedy if it turned out that morality were not so cleanly structured as to have one".[3]

Wolf thinks that Kantianism, contractualism and consequentialism all capture something important about value and about how to lead and make sense of ethical life. Yet she worries that there is deep tension and disagreement between these theories, and that this is inevitable since what they say is of value and the way in which they capture the valuable differs, often fundamentally. Attempting to reconcile these theories will result in a dilution of their individual visions of what the ethical life is. Involved in this, I take it, is a worry that we may well lose some aspects of our ethical life that each may show us to be valuable, and that we may lose an appreciation, in part or whole, of why they are valuable.

Wolf's main example in this regard concerns autonomy and consent. She focuses on the tension between Kantianism and consequentialism, and specifically the tension that seemingly exists between a concern to respect autonomy and a concern to produce optimific results. In short, Wolf notes that Parfit's commitment to an objective, value-based account of reasons means that what many take to be

important and morally significant about consent drops out of the picture. Under Parfit's construal, she thinks, when we think about whether to act in a way that will affect some person, we think not about how she has consented, or how she would consent were she able to, but only about the reasons that relate to our action, reasons that justify choices that she herself *could* (but not necessarily *would*) endorse.

This idea conflicts with the value of consent given by many theorists, including Kant.[4] As I read Wolf, it also conflicts with a prevalent, everyday understanding of consent. The idea is simply this. We may well be able to maximize best outcomes if we φ, but a (central) person in the situation has not consented to our φ-ing, or has expressly forbidden us to φ, or (we can reasonably imagine) would refuse to consent to our φ-ing if asked. Such refusals stand as important checks on our action. If we do decide to φ, then we would be overriding what this person has said she wants to happen, or would say if asked. According to Wolf, Parfit's treatment of consent, with its direct and explicit link to the reasons that exist, allows him to introduce a concern for optimific results and drop respect for what people choose and would choose. Wolf illustrates this with the example of *Means*. She says that there are things that count in both directions in this case, but that it would be odd to say that in saving White's life one had satisfied some Consent Principle.[5] The point is that there is no real acknowledgement of the respect that we should give to Grey, or anyone, in the exercise of their own practical reason.

At the end of her section 'Consent', Wolf says:

> The problem with [Parfit's] suggestion, as I have argued, is that it leaves what may be considered the moral point behind a consent principle behind. It leaves consent behind, and the respect for autonomy, from which the value of consent might be thought to derive. If one is concerned in the first instance not in formulating a supreme or decisive principle, but rather in registering and articulating important (but possibly competing) moral considerations, the need for unanimity would not be allowed to transform one's principles in this way.[6]

Consent drops out for Parfit, according to Wolf, but what is really interesting is why it does so. It drops out because of Parfit's aim to formulate a supreme principle involving as it does – to pick a label – the *smoothing over* of seemingly profound moral differences.[7]

We can push further. There is a feature of Wolf's criticism and Parfit's reply that reveals the nub of their discussion. At one point Wolf discusses a 'trolley' case. We are to imagine being in a position to push a man onto the tracks to stop some runaway trolley and hence save people on the track.[8] She claims that people are resistant to pushing the man not only because he is innocent. What is also involved – indeed what is "distressing" – is that someone else is deciding what to do with someone's life, even if many other lives could be saved as a result. Much of the appeal of autonomy lies in choosing what to do with your own life, where it is you who is "calling the shots", and Wolf imagines that we can project such a

view onto other people and imagine what it is like for them. This is not a mere preference, for Wolf, as opposed to a value (her contrast). It should be classified as more important than that. She argues that this preference is something that everyone could adopt and that we should treat such a preference as rational. So, for example, it is perfectly rational to accept a principle that favours leaving some man on a bridge (where if we pushed we could save many) and prefer it to a principle that says we should push. She says:

> If it be granted, therefore, that a person may rationally prefer to maintain immediate control over his body and his life to minimizing his risk of loss of life and limb, then Parfit's argument that Kantian Contractualists must support a form of Rule Consequentialism will not go through. Even if we grant Parfit's claim that everyone could rationally accept optimific principles, as I am happy to do, we would also have to admit that everyone could rationally accept nonoptimific principles, in particular principles which would more strongly protect people against interference from others in the control of their own bodies.[9]

This continues her discussion, for in effect she draws a distinction between a preference for welfare and a preference for autonomy, and adds that some Kantians or Kantian Contractualists would further claim that preference for autonomy over welfare would be "uniquely rational". To Wolf's mind, the value of autonomy is 'irreducibly important' for some people and this is something Parfit fails to recognize. Given her main theme is to emphasize the complexity and variety of ethical life, we can readily class this as just one example of a difficult or impossible choice amongst many.

Parfit's response to this passage is revealing.[10] He casts Wolf as saying that "everyone could rationally choose that everyone accepts some such principle even though this principle would not be optimific" and says that both claims could not be true. Why not? When, as Kantian Contractualists, we ask which principles everyone could rationally choose, we presume they know all of the reason-giving facts. If these autonomy-protecting principles were not optimific, then they simply would not be chosen: people would have clear impartial reason to refrain from choosing them. In effect, Parfit sees no possibility of a clash between a rational preference for, or a valuing of, autonomy and a concern to adopt optimific principles. To rationally prefer some principle *simply is* to see it as an optimific principle, and vice versa.

Parfit goes further in sorting out Wolf's criticism by distinguishing welfare from optimificality.[11] Wolf seems to treat the two as synonyms, but Parfit is at pains to make clear that he is not committed (and is indeed not trying) to further a welfarist account of rule consequentialism. He is concerned only with those principles that make "things go best",[12] and, as he sees matters, if we had a situation where everyone rationally chose that everyone accept some autonomy-protecting principle, then, again, this is simply what it would be for everyone to accept the

principles that are optimific, since such principles would make things go best in the impartial-reason-implying sense.

There is here a fundamental and revealing misunderstanding on Parfit's part. He uses 'rational' to indicate that one is responsive to reasons in an impartial sense. When Wolf uses 'rational', I believe she has a different, perfectly acceptable, conception of 'rational' in mind. When she imagines that people might prefer autonomy over something that could maximize welfare, she has in mind that it would be perfectly (morally) reasonable and understandable that people would make such a choice, for the reasons she gives, such as wanting to call the shots. (And not only *could* people have such a reasonable preference, but they *do*.) We can also note the connected point that whilst Parfit is right to indicate that he is not advancing a welfarist conception of 'best', his notion of 'best' does reveal a blind spot. He imagines that if everyone rationally chose for everyone to accept autonomy-protecting principles, then this is just what it would be to choose the best principles: that is, optimific principles. His notions of 'best' and 'optimific' are 'singular' notions, at least when discussing the Triple Theory. (More on that important qualifier in the next section.) Throughout much of *OWM* he is searching for the set of principles and moral ideas that tell us what we should do, seemingly singularly and uniquely (or, at least, for an overarching principle that helps us to choose which substantive principles should guide our actions).[13] Wolf, on the other hand, wants to contrast choosing autonomy over welfare or other notions, whereby in some cases there is no obvious single solution and some cases that have significant moral 'residue'. This will mean there are tensions between various notions, some of which will be impossible to resolve, leaving us to render the value of such notions irreducible with respect to other values and things valued. In a particularly complex and fraught situation, it can be reasonable and rational to choose to protect autonomy but also reasonable and rational to choose to promote welfare, or keep one's promises, or whatever. A normative ethic can be decent and good whilst leaving such a tension in place. To always wish for some singular best set of principles to guide us is to distort the character of the ethical life. As Wolf says:

> For Parfit, appreciation of the different evaluative outlooks poses a challenge which he aims in this book to meet: to unify, systematize or otherwise combine the insights gleaned from these perspectives to reach a single, coherent moral view that can guide our actions in a way that is free from moral remainders and normative tensions. Though I think I understand the wish to reconcile the different traditions and transform their ideas into a single, unified whole, I am less gripped by it than many other philosophers.[14]

Wolf is not just emphasizing the complexity and variety of our moral lives, nor just that ethical theories perceive and cast what is valuable in different ways. Many people – including Parfit – can agree with that. She is arguing that many

such theories are reasonable to hold in some sense because they each have some grain of truth in them. This is so since the ethical life contains many values, many of which are incommensurable. The word 'incommensurable' makes only one appearance in the whole of OWM, in Wolf's conclusion.[15] Parfit never uses it. But as far as Wolf is concerned, it is crucial to his account that values and things held to be valuable are commensurable.

Having now brought up the notion of incommensurability, we should begin to nail those three themes I introduced at the start of this section.

Three interrelated themes

Despite not using the word 'incommensurability' in OWM, Parfit does discuss matters pertinent to it.[16]

In Chapter 34, 'Agreement', he considers those that seek to attack moral realism or moral truth on the basis that there is a significant amount of moral disagreement. To cut a long story short, he runs through various features that may explain why there is so much disagreement and which in turn do not threaten the possibility of there being moral truth. As well as expected examples, such as ignorance of or disagreement about nonmoral facts, he also says:

> Some other moral disagreements are not about *which* acts are wrong but *why* these acts are wrong, or what *makes* them wrong. Different answers are given by different systematic theories, such as those developed by Kantians, Contractualists, and Consequentialists. Such disagreements do not directly challenge the view that we are able to recognize some moral truths . . . . [W]e would expect there to be more disagreement about these other questions [as to why acts are wrong]. As I have also argued, however, when the most plausible systematic theories are developed further, as they need to be, these theories cease to conflict.[17]

This passage occurs in §121, 'The Convergence Claim'. Whilst it clearly chimes with Parts Two and Three and therefore allows us to sustain Wolf's criticism, other parts of §121 may cut against the overall picture.[18]

I have floated the idea that Parfit thinks there should be a singular and unique sense of best, with which we could decide each problem. There is evidence for that in Parts Two and Three, but notice that I qualified this claim in the previous section. In §121 Parfit also discusses comparisons and makes two relevant points. First, he talks of precision. He thinks that it is a mistake to assume a universal linear model of best to worst for all judgements. Sometimes it is impossible to compare apples with oranges, or a job in one city with a different job in another city, and say definitively which one is best. Hence there may be apparent disagreement: a clash of views that, on Parfit's view of disagreement, does not constitute a (proper) disagreement. We can say in such cases that each of the things we are comparing is equally good where we mean 'imprecisely equally good' but not rank

I suggest that the ordinary normative concept of wrongness can be character-ized as the concept (roughly) of *the property of being a violation of an important, authoritative moral standard where blame is warranted, other things being equal, when a person violates such a standard.*[47] Of course, the concept of an authoritative moral standard is also normative, but it should be no surprise that a characterization of the normative concept of wrongness is normative. The nonanalytic naturalist agrees with the non-naturalist that normative concepts are not analyzable in non-normative terms, but she maintains that the properties picked out by the normative concepts are natural properties.

That is, the naturalist holds that if there is a property represented by the ordi-nary concept of wrongness, it is a natural property. To explicate the nature of this property, a normative naturalist presumably would need to provide a naturalistic account of authoritative moral standards. I have attempted to do this elsewhere with my theory of pluralist-teleology, which I sketched earlier in the chapter.[48] I say that the authoritative moral standards are, roughly, those with the currency in society that would enable the society to ameliorate the "problem of sociality."[49] Of course, I am not claiming that this is a conceptual truth or that a person with the ordinary concept of wrongness must have this view about the nature of authoritative standards. The normative concept of wrongness is simply the con-cept (inter alia) of being a violation of such a standard.

Recall, as I pointed out above, in the section "Further responses: the role of the normative concepts," that if normative-judgment internalism is true, the nature of normative concepts, plausibly, is such as to explain the immediacy of moral motivation for people who have normative beliefs. But even if normative-judgment internalism is *not* true, there may be a way of thinking of wrongness, for instance, such that, if one thinks of wrongness in *this* way, and if one believes that, say, lying is wrong, one will be motivated accordingly, at least to some degree. Call this the *internal* way of thinking of wrongness. This is the idea that I now wish to explain. I will distinguish between the ordinary normative concept of wrongness and the motivationally laden internal *way of thinking* of wrongness. As I will explain, an ordinary morally motivated person would typically have the internal WOT of wrongness, for she would be motivated to avoid wrongdoing, and she would think of wrongdoing in a way that is 'colored' by her motivation in a way I will explain.[50] Such a person would have the ordinary normative concept of wrongness, but, as we might say, she also would think of wrongness as normative *for her*.

Consider the difference between the way that a person who loves cats may think of cats and the way a person who is neutral in his feelings about cats pre-sumably thinks of cats. Both have the ordinary concept of cats as *animals with the cat properties*, we might say. However, the cat-lover I have in mind thinks of cats as animals that are lovable *because* they have the cat properties. That is, she thinks of cats as *animals that deserve to be loved in the way that I love them because they have the cat properties.* The cat-lover is attracted to cats because of her love, and *in being attracted*, she thinks of them as *lovable*. Obviously, the cat-lover must

44