Derek Parfit, ON WHAT MATTERS: VOLUME ONE and ON WHAT MATTERS: VOLUME TWO

T.J. Mawson

On What Matters: Volume One and *On What Matters: Volume Two*, by Derek Parfit. Oxford: Oxford University Press, 2011. 540 pages and 825 pages. $35.00.

T. J. MAWSON, Oxford University

This is a substantial work, running to two densely-argued volumes. Each volume is divided into three parts. Volume one lays out the arguments for Parfit's particular theory of ethical normativity—what it is, according to him, that makes actions right and wrong in an ethical sense. Parfit argues for what he calls "The Triple Theory," which is a version of what is sometimes called "Ideal Rule Consequentialism." We should follow those principles the universal acceptance of which everyone could rationally will; nobody could rationally reject; and these principles are those that would be optimific. The material here is an expanded version of that given in his Tanner Lectures, as delivered at the University of California in 2002. A condition on the lecturer delivering these lectures is, I am told, that he or she ensure that the published version be accompanied by a discussion with commentators. So it is that the first part of volume two, part four of the whole then, starts with four commentaries on (critiques of) Parfit's views as they've been laid out in volume one. These are given by Thomas Scanlon, Susan Wolf, Allen Wood, and Barbara Herman. Part five, Parfit's responses to these critics, follows swiftly. In total, these two parts of the second volume add up to a mere 260 or so pages (one can say "mere" of this figure in the context of this work). The rest of volume two is taken up with part six. Here Parfit takes a step back, as it were, by discussing, not the pros and cons of various theories of ethical normativity, but the metaethical issue of the status of this normativity. Parfit argues for a "Non-metaphysical, Non-naturalist, Cognitivism" about reasons for action; claims about these sorts of reasons can be true, but their truth is a "non-ontological" truth, one which we come to know with a special ability—akin to mathematical insight.

As regards the ordering of the whole, I would have preferred the discussion of the more basic—metaethical—material that is found in part six before anything else or at least in with the material on subjectivism/objectivism that is in part one of volume one; then the laying out of the particular normative theory that is the subject matter of the second and third parts of the first volume as it stands; and only then any commentaries/critiques and responses to them and any that there have been to have taken in all of what would then have been the foregoing material. Be that as it may, as well as a contents page (as one would expect), each volume comes with a detailed summary of what is going to be argued for and where it's going to be argued and a comprehensive index. These two features are particularly helpful to the reader in navigating the work over its several hundreds of pages. Even the most careful and diligent would find it, I think, hard to follow the argument of the whole without refreshing his

or her memory now and again by looking back to sections the arguments and conclusions of which are being referred to or assumed in what follows them and these features certainly facilitate that.

Though large numbers of moral philosophers, probably the majority, would have their most fundamental disagreements with Parfit at the meta-ethical level—i.e., they would take issue with some of the claims he makes in parts one and six—as I am broadly sympathetic to the views outlined and argued for there (and indeed to the manner of the argumentation), in what space remains to me, I shall treat—I fear it will need to be in very broad-brush terms—of the material of parts two and three of volume one.

Parfit starts volume one by paying homage to his two greatest heroes, Kant and Sidgwick, and it is in their shared tradition that he firmly places himself in at least one respect: he is in the business of searching for a "supreme principle of morality." One does not find a mention of Moral Particularism in the index to either of Parfit's volumes; Jonathan Dancy is thanked at one stage, but his views on this front do not seem to have made an impact. Rather, just as one might—at least if unaffected by the "Disunity of Science" movement—search for a Theory of Everything in the Natural Sciences, so Parfit searches for one in the Moral Sciences. (He is, to my mind, rather too brief in his response to Wolf on a similar point (volume two, page 155), saying simply that it would not be in itself a tragedy were the search for a single supreme principle of morality not to succeed.) Whilst natural scientists paradigmatically—though of course not exclusively—conduct real experiments to further their investigations and whilst those influenced by so-called "Experimental Philosophy" (again, interestingly, no mention of this in the index of either volume) have conducted experiments of a similar sort with a view to illuminating the questions Parfit is investigating, Parfit proceeds by using thought experiments in the classic analytic-philosophical way. A prima facie plausible moral principle of some generality is articulated; a scenario is imagined whereby the principle would clearly dictate that the agent do (or not do) a particular thing; it is expected of the reader that he or she will have an intuition about what the agent should do (or not do) in the imagined scenario which differs from this; and thus it is supposed that a counter-example has been given to the original principle—a conclusive refutation of a bold conjecture, as one might say. It is hoped one will find oneself saying things like, "You know what? You're right. I *would* think it right to utilize the remote control device so as to launch a fat man from a bridge into the path of an oncoming train were I to be possessed of preternatural certain knowledge that doing so was the only way of preventing that train killing five. So I now know that the prima facie plausible principle that it's always wrong to treat other people merely as means can't be right." For this method to be taken to have resulted in progress, obviously one must have epistemic confidence in such intuitions in such contexts sufficient that those principles with which they conflict and which might otherwise have been attractive to one can reasonably be modified/abandoned. And one's

doubts about the reliability of one's intuitions in these contexts cannot but be heightened by reflecting on the fact that the conditions specified in the thought experiments eliciting them are often very unlikely to be realized; sometimes even, they cannot be realized. Why should we have much confidence in our moral intuitions about such bizarre circumstances? "If pigs could fly, the price of pork would decrease" seems intuitively plausible to me, but I can imagine one of my colleagues in Economics convincing me otherwise from prima facie plausible economic principles and my—on reflective-equilibrium balance—rejecting the intuition, not the principle. Parfit is of course aware of this; indeed, he makes the same point himself towards the end of volume one. But one doesn't need to be as suspicious as is one of his commentators—Wood, who goes to some lengths in making the case that such "trolley" problems are "worse than useless for moral philosophy" (volume two, p. 68)—to appreciate that, given this methodology, there is a limit to what Parfit can do to ensure that people who simply don't share his intuitions or level of confidence in them do not alight from his train of reasoning before it finally pulls into the Triple Theory station. Be that as it may, the Triple Theory is his final destination and, as such, it deserves some scrutiny: "An act is wrong just when such acts are disallowed by some principle that is optimific, uniquely universally willable, and not reasonably rejectable" (volume one, p. 413).

Although the name and this summary of the Triple Theory may suggest three criteria for wrongness, criteria which are each separately necessary, and only jointly sufficient, for moral wrongness, in fact, Parfit argues, these "three" criteria are equivalent: no act satisfies one without satisfying the other two. We all ought to follow those principles which everyone could rationally will to be universally accepted; and these principles can be seen to be those that—were they to be universally accepted—would lead to the best overall consequences, judged from an impersonal perspective. From an impartial point of view, a sort-of Rawlsian original position, these are the principles which everyone would find themselves having most reason to want to be universally adopted; nobody could reasonably reject them. We can assume that there is some set of principles such that its universal acceptance would make things go better than the acceptance of any other set (or thinking like Act Consequentialists) and this set of principles is the one which we should follow. So much for the impartial perspective, but what of things as seen from my partial perspective? Might I not have some outweighing reasons of a self-interested sort to follow less burdensome principles than these optimific ones, or perhaps to abandon principles altogether and be an amoralist? No, according to Parfit: the goods which come to one—including the good of acting partially for those that one cares about (one's friends and family, for instance)—would in fact be achieved best for one by following these principles; the sort of Rule Consequentialism to which we are led does not fail by virtue of the over-demandingness objection or due to integrity or alienation concerns. Or so he argues.

In advancing his case for the Triple Theory, Parfit persistently returns to the analogy of climbing a mountain from different sides; and one of his primary aims (the introduction states it as *the* primary aim of parts two and three of the first volume) is to show that, now that we—with Parfit—are in sight of the peak of this mountain, we can recognize the routes taken by what are usually thought of as differing schools of normative theory as routes which ultimately come together in endorsing the Triple Theory. The three schools of thought that Parfit aims to unify are Consequentialism, Kantianism, and—not, as one might have expected to feature as the third member of a trinity of normative theories, Virtue Theory, but—Contractualism. (Perhaps Virtue Theory could be swept up into the converging consensus too; perhaps further thought experiments could establish that we intuitively judge virtuous all and only those who habitually act in accord with the Triple Theory.) Volume one thus spends some time making the central notions and arguments of these three schools of thought as clear and defensible as possible, if needs be by saying things which the progenitors and archetypal adherents of the schools—e.g., the historical Kant—would have eschewed (or even explicitly *did* eschew). So be it; Parfit's business is, after all, to get at the truth about moral reality. In this business, just as he would be a fool not to seek to learn from the great philosophers of the past, so he would be an equal fool slavishly to follow them down what we can now see, with the aid of the right thought experiments, are obvious dead-ends. After Parfit has argued that the three routes come together in this manner at the top of the mountain, we are thus left with the thought that, as convergence is a mark of truth, the convergence of these three great philosophical traditions means we are truly at the top of the mountain and we were right, after all, to think of the top as having a unique peak, a single principle of morality. (N.B. although this is certainly the thought with which Parfit closes volume one, Parfit actually uses the name "The Convergence Argument" for another point, made earlier in that volume [that Kantian and Scanlonian Contractualism can be combined].)

It would of course be possible to complain that Parfit has "cherry picked" rather from these traditions and that this cherry-picking devalues his "discovery" of convergence to at least some extent. Parfit has certainly drawn out from these three traditions the strands of thought that have commonalities with one another, leaving behind the strands that do not. Those strands that he has left behind are then, uncontroversially, the most distinctive of the traditions, and it's arguable that they are more central to the traditions than the strands that he's taken up to weave together into his Triple Theory. But it's arguable that they're not. What's certain is that, depending on one's judgment on this issue, the convergence Parfit claims will seem more or less significant—more or less a discovery, rather than an artifact—and thus it will seem to speak in favour of the Triple Theory to a greater or lesser extent accordingly, in favour of the Triple Theory and in favour of the presumptions behind it. At this stage, one might consider the

material of part six, which, as I have indicated, seems to me to have methodological priority to the rest, and part one. Proponents of these three traditions would agree in large part on the material of these parts, but many moral philosophers would not, thus making any convergence between the three traditions that is discovered a convergence between a proper subset of moral philosophies, not the totality of moral philosophies (albeit that Parfit would criticize those "moral philosophies" that fell outside this subset as being all variants of nihilism and thus perhaps not really in the set at all).

Let me finally turn to Parfit's most detailed statement of the Triple Theory and close by stating, very briefly, an (inconclusive) objection to it.

> An act is wrong if and only if, or just when, such acts are disallowed by some principle that is
>
> (1) one of the principles whose being universal laws would make things go best,
>
> (2) one of the only principles whose being universal laws everyone could rationally will, and
>
> (3) a principle that no one could reasonably reject (volume 1, 412–413.)

Of course one cannot show this theory to be wrong by running it up against the intuitions produced in response to just one scenario, but it does seem to me that—as with other variants of Ideal Rule Consequentialism—the Triple Theory is subject to the criticism of licensing, indeed requiring, an implausible "Utopianism," and I want to close by drawing attention to this. (I must mention that Parfit is alert to this concern, which he calls "The Ideal World Objection," and he suggests ways in which it might be avoided (volume one, p. 312 ff.), ways to which I cannot do justice here; I can only, rather weakly, state my view that the Utopianism/Ideal World Objection survives them.)

By way of giving the flavor of the objection, consider the laudable intentions behind the creation of Esperanto, an easy-to-learn and politically neutral language, the founder of which hoped would foster world peace and understanding. Who can doubt that were everyone to be following those principles that would make things go best, they'd learn Esperanto? And who can doubt that, even as it is, *ceteris paribus*, the more people who learn to speak Esperanto, the better the world will be? Well, I can't. So, was I acting wrongly a moment ago when I willfully failed to log-on to the free online Esperanto classes to which Google directed me when I looked up "Esperanto" to check on my facts? It seems as if the Triple Theory would have it that I was; that I act wrongly in not learning Esperanto (given that I have plenty of spare time in which I could learn it; the learning of it wouldn't displace other even-more-worthwhile activities from my life, just other in-themselves-worthless hobbies/pastimes). However, whilst I don't doubt that, ceteris paribus, the world is better the more people who know Esperanto, I have a pretty strong intuition that I'm not acting

wrongly by willfully failing to learn Esperanto. And it's an intuition that springs up in response not to some bizarre thought experiment, but in response to a situation that I am in fact in; I have just navigated away from the webpage that would have given me access to the free classes and I don't intend to return to it. Given that we don't actually live in a world of saints, shouldn't we behave differently from the way that we should behave if we did? Aren't we at least permitted to behave differently? It seems to me that we should and that we are so permitted; it seems to me that according to the Triple Theory, we should not and are not. And thus the Triple Theory seems wrong to me.

On What Matters is a closely-argued and detailed investigation of some of the main metaethical issues (there are notable absences: discussion of God and free will is rather brisk) and of the commonalities between three great traditions of normative theory: Kantianism, Consequentialism, and Contractualism. I hazard that nobody will agree with it in its entirety, but anyone with interests in any of the topics on which it dwells will find much of value in it and that it amply rewards what it demands: careful reading and re-reading.

In Adam's Fall: A Meditation on the Christian Doctrine of Original Sin, by Ian A. McFarland. Wiley-Blackwell, 2010. 256 pages. $120 (hardcover).

MATTHEW BARRETT, California Baptist University

Few doctrines are as difficult and controversial as the doctrine of original sin. To the modern thinker, the doctrine of original sin is absurd. How can one person's sin eons and eons ago determine the status and nature of all humanity? But when we open the pages of Scripture, we quickly see that the doctrine of original sin is a biblical doctrine, taught by the apostle Paul himself. Therefore, it is safe to say that the doctrine of original sin is a Christian belief. Where controversy arises, however, is in exactly how we, as Christians, should define original sin.

Ian McFarland has sought to tackle this very issue, though from a more philosophical and historical perspective. His thesis is straightforward and ambitious:

> In reaction to a wide range of criticisms leveled against the idea of original sin, a number of Christian theologians in the modern period have attempted to develop a doctrine of sin in which the idea of original sin is heavily qualified or even rejected. Against these perspectives, I will argue that it is not only theologically defensible, but inseparable from the confession of Jesus Christ as Savior and Lord. Indeed, I will defend the doctrine in what is arguably its most extreme form, as developed by Augustine and later defended in the Reformed theological tradition under the designation of "total depravity"—the claim that no aspect of our humanity is untouched by sin. (x)