On What Matters, by Derek Parfit. Oxford: Oxford University Press, 2011. Volume One, pp. xlviii + 540. Volume Two, pp. xiv + 825. H/b £30.00.

Written in his trademark prose — repetitive, intricate, and rhythmic — *On What Matters* is, among other things, the published version of Derek Parfit's 2002 Tanner lectures, 'What We Could Rationally Will'. It is a major philosophical event: the second monograph of one of the most important and influential living philosophers. And its scale and ambition are massive. This book presents a comprehensive theory of the metaphysics, epistemology, and substance of ethical thought. Its originality is often striking and its arguments profound. *On What Matters* is a monument that will shape the field for many years.

Although it goes far beyond the original lectures, this book preserves their informality, in part. Its style is engagingly, often eccentrically, conversational, filled with memorable examples and turns of phrase, passionate, at times inspiring, and enjoyable to read. The book begins with a helpful introduction by Samuel Scheffler, and a lovely preface by Parfit, gently mocking his 'two masters', Kant and Sidgwick. There is also a surprisingly personal chapter on Nietzsche's immoralism.

At the heart of Parfit's book is an argument for convergence in moral theory, according to which the most plausible version of Kantian ethics is a form of contractualism that is not only consistent with, but in fact entails, the most plausible version of consequentialism. Disagreement in moral theory is, or should be, less extensive than is often supposed. It is therefore less threatening to moral belief. In the background of this argument is a non-reductive theory of practical reason developed at the end of the book. The rest of Volume Two is given to discussion of Volume One, with responses by Susan Wolf, Allen Wood, Barbara Herman, and T. M. Scanlon, along with Parfit's replies. In my view, the first volume will have a deeper and more positive impact on moral philosophy than the second. Its claims are more distinctive, its arguments are stronger, and there is more to learn from them. Partly for these reasons, partly because there is more in these volumes than a review — even a long one — could hope to address, I will focus mainly on the convergence argument. But I will say something, first, about the metaphysics of reasons.

Parfit defends a form of Non-Naturalist Cognitivism, according to which we know irreducibly normative truths. This defence has four parts: an account of what he calls 'Externalism' about reasons; a series of arguments against reductive Naturalism; a critique of Non-Cognitivism; and a response to epistemological problems raised by the Non-Naturalist view. Parfit's Externalist insists that 'reason' has a practical, normative sense that cannot be analyzed in non-normative terms or in terms of procedural rationality (Vol. 2, p. 270). Externalists thus reject Analytical Internalism, on which the concept of a reason is explained in normative procedural terms, and

Analytical Naturalism. It is an odd way to define 'Externalism', since it allows Non-Analytical Internalists to be Externalists, and since Analytical Naturalists, who cannot be Externalists in Parfit's sense, may also reject Internalism. Parfit's view is in any case simpler: that 'reason', 'ought', and 'should' are conceptually or analytically irreducible. The senses of these words that matter to ethics cannot be analyzed in other terms.

This leaves room for reductive Naturalism, on which normative facts can be stated with non-normative concepts, so long as the relevant identities are not analytic truths. Although the practical 'should' does not express the concept of what would satisfy one's final desires, the suggestion might go, that is just what it is for an act to be what one should do. Parfit takes this kind of position seriously and argues against it. But his arguments are likely to frustrate. According to the Normativity Objection, natural facts, ones that can be stated using only non-normative concepts, could not be normative (Vol. 2, pp. 324–7). We know this in much the way that we know, *a priori*, that justice could not be the number four. If Parfit is right about this, reductive Naturalism fails. But, as he notes, the Normativity Objection is a mere expression of belief. Taken as an argument, it begs the question. Parfit's opponents will contend that Naturalism is more akin to the identification of conscious states with physical ones: controversial, but by no means obviously wrong. The Normativity Objection gives no independent reason to reject their claim.

Parfit suggests a more substantial argument against ethical Naturalism, though its details are hard to make out. He presents what I read as a single line of thought in three distinct steps: the Fact-Stating Argument, the Triviality Objection, and the Lost Property Problem (Vol. 2, pp. 334–56). Putting the pieces together, we find a challenge to the reductive Naturalist to explain the informational content of the alleged discovery that an act's being what one should do is its being such as to satisfy one's final desires. Since we are dealing with Non-Analytical Naturalism, this challenge may seem weak. The identity is not an analytic truth or a concealed tautology, so why should it fail to be informative? The answer is that, according to Parfit, if it is informative to learn that what would satisfy one's final desires is what one should do, this 'information must be statable […] as the claim that such acts [ones that would satisfy those desires] would have one or more other, different properties' (Vol. 2, p. 344). Compare the discovery that heat is molecular kinetic energy. This can be informative because the concept *heat* is the concept of the property that turns solids into liquids, liquids into gases, causes sensation of warmth, and so on (Vol. 2, pp. 334–5). If we are reductive Naturalists, we must analyze the concept *should*, as we analyze the concept *heat*, in terms of other properties. But then we face a dilemma. If the concepts in terms of which we analyze *should* are entirely non-normative, we are Analytical Naturalists after all. And if we appeal to a normative concept, we are no further on. Naturalism must apply to this concept too, and the same dilemma will arise. Assuming that circularity is out — that we cannot analyze normative concepts in terms of each other, or in some infinite series — we must at some point give the first response, which appeals to non-normative concepts. The upshot is that, unless we are Analytical Naturalists, we cannot be Naturalists at all.

If this is Parfit's argument, it rests on a contentious claim about the nature of concepts. Parfit assumes, in effect, that when two concepts refer to a single property, and the first is unanalyzed, the content of the second can be explained in other terms. The challenge to the Naturalist is then to analyze normative concepts in a way that reveals the informational content of the alleged identity. But the Naturalist should refuse this demand. She should deny that the concept *should* can be analyzed in terms of other concepts. It is informative to learn that what would satisfy one's final desires is what one should do, but not because one discovers that doing what satisfies one's final desires has a property distinct from that of being what one should do. One learns, precisely, that it is what one should do, information that cannot be expressed in any other way. If the mode of presentation of the concept *should* is non-descriptive but distinct from the mode of presentation involved in the concept of desire-satisfaction, the relevant identity can be informative while avoiding the dilemma above.

Assessing this view would take us deeper into the philosophy of language than Parfit's book descends. It is notable that, while he devotes five valuable chapters to the epistemology of Non-Naturalism, Parfit does not ask how, or in virtue of what, our normative concepts attach to the irreducible properties they do. One way to make this question vivid is to suppose that there are many non-natural properties that share the structure of reasons in relating considerations to acts and agents, but are extensionally different. Is this scenario possible? Nothing in Parfit's book suggests otherwise. But if it is possible, we can ask: why does 'reason' attach to the reason-relation, not to one of its competitors?

Supposing this puzzle can be solved, there is another to confront. If it is possible to guide one's life by these other relations — and again, nothing in Parfit's book suggests otherwise — we can imagine people who do so. We respond to and talk about reasons, while the members of some other community respond to and talk about reasons*. How should we react to this? Of course, we can tell ourselves that we are right to do what we are doing, in that we *should* respond to reasons. But the members of the other community use 'should' for a concept that stands to reasons* as *should* stands to reasons. While they agree with us that they are not acting as they should in *our* sense of 'should', they say to themselves, correctly, 'Still, we are doing what we should'. It is hard to believe that there is no feature of the concept *should* to break this symmetry, that there is nothing to say about the content of this concept and how it differs from theirs, beyond the fact that it stands for what one should do and relates to reasons, not reasons*. Can that be the story's end?

Let us go back to the beginning. Though Parfit thinks otherwise, little of significance in Volume One rests on the truth of Non-Naturalism or the arguments of Volume Two. The assumptions Parfit needs are about the content, not the metaphysics, of our 'non-deontic' reasons, reasons that do not consist in an action's being morally wrong. According to Parfit's 'wide value-based objective view', when one of two acts would benefit strangers, while the other would benefit us or those with whom we have close ties, we often have sufficient non-deontic reason to act in either way (Vol. 1, p. 137).

It is in this context that Parfit explores the modified version of Kant's Formula of Universal Law that he calls 'Kantian Contractualism'. According to this theory, we ought to follow principles whose universal acceptance everyone has sufficient non-deontic reason to will (Vol. 1, p. 342). Unlike Kant's formula, Kantian Contractualism makes no appeal to the problematic notion of a 'maxim' (Vol. 1, pp. 289–98). Nor does it ask only what *I* could will. If we restrict our focus in that way, we permit acts of exploitation whose agents would be unharmed by their general performance. As Parfit argues, those with power might have sufficient non-deontic reason to will principles that take advantage of others, knowing that there is little risk that they will be exploited in turn (Vol. 1, p. 334). The Kantian Contractualist looks instead for principles that *everyone* could rationally will, even those who are most vulnerable to their bad effects. It might seem that nothing could pass so strict a test. How could the principles I have sufficient non-deontic reason to will, given my particular interests and desires, overlap with the principles you have reason to will, if your desires and interests are different? This question would have no answer on many accounts of practical reason. If reasons turn on what we want, or if we have non-deontic reason only to promote our own well-being, there will be no set of principles that everyone could will. Things look different on the wide value-based objective view. On this account, we often have impartial non-deontic reasons, and we lack decisive reason to favour ourselves. If that is right, there might be principles whose universal acceptance everyone could rationally will.

Though Parfit never presents a direct argument for Kantian Contractualism, he approaches it through a brilliant discussion of alternative views, including Kant's Formula of Humanity, the Golden Rule, and versions of contractualism that owe more to Hobbes and Rawls. These chapters are terrific: the most perceptive, enlightening introduction to moral theory I know. I say 'introduction' not because the treatment is simplistic — on the contrary, it is extraordinarily subtle — but because it is so clearly written, so carefully argued, and so sharply focused on the most essential points. Parfit argues convincingly that Kantian Contractualism is the most plausible contractualist view, that it shares the most compelling features of its rivals, and that it lacks their most decisive flaws. Along the way, he makes countless original and constructive moves. To give just one example, it is a problem for contractualism in most of its forms, but also for such theories as Rule Consequentialism, that they permit us to act on principles that work well under full compliance but are disastrous otherwise. Parfit considers the policy of never using violence unless others do, in which case one tries to kill as many as one can (Vol. 1, p. 315). As he points out, we might have sufficient non-deontic reason to will that everyone act on this conditional principle, since that would lead to a world without violence; but acting on it in the actual world would be absurdly wrong. Parfit's solution is to look for principles that we could rationally will any number of people to act on, in that '*whatever* the number of people who *don't* act on this [principle], *everyone else* does' (Vol. 1, p. 318). Such principles would take account of what others are doing so as not to go wrong in conditions of partial compliance.

We might worry that, while such principles would not go wrong, there are no such principles: the new requirement is too difficult to meet. Since Parfit omits the qualification from the usual statement of his views, and from his central arguments, I will ignore it here. According to the crucial argument of Volume One, we can reconcile Kantian Contractualism with Rule Consequentialism by showing that the first entails the second. The argument runs as follows. Let us call an outcome 'optimific' when it is the outcome we have most reason to want from an impartial point of view. Suppose there are some principles whose universal acceptance would be optimific. According to Kantian Contractualism, we ought to act on principles whose universal acceptance everyone has sufficient non-deontic reason to will. Everyone would have strong impartial reasons to will the universal acceptance of the optimific principles. What is more, Parfit argues, no-one would have decisive non-deontic reason not to will these principles. Nor are there any other principles whose universal acceptance everyone would have sufficient non-deontic reason to will. Conclusion: the optimific principles are the only principles whose universal acceptance everyone would have sufficient non-deontic reason to will. If Kantian Contractualism is true, so is Rule Consequentialism. These theories do not conflict.

Parfit goes on to argue for a further convergence, with Scanlon's Contractualism, according to which we should follow principles no-one could reasonably reject (Vol. 1, pp. 411–12; see also Vol. 2, pp. 191–259). Though it might be disputed, this argument is less striking than the alleged convergence of Kantian ethics and consequentialism. What should we make of this surprising claim?

Parfit spends most time defending one premise of his argument: that no-one would have decisive non-deontic reason not to will the optimific principles. He does so by asking what such reasons might be. For instance, in Lifeboat, I am stranded on one rock and five people are stranded on another (Vol. 1, pp. 380–2). An optimific principle would require you to save them, not me. But, it might be argued, if some other principle would have you save me instead — perhaps I am on the nearest rock and the Nearness Principle requires us to save the nearest group — I would have decisive non-deontic reason to will this principle, even though it is not optimific. Parfit denies this claim about non-deontic reasons. But, he argues, even if it were true — even if I had decisive reason to prefer that *you* accept the Nearness Principle, saving my life at the cost of five — I would not have decisive reason to will the *universal* acceptance of this principle, at the cost of many. If everyone accepted the Nearness Principle, instead of saving the greater number, millions of lives would be lost. On any plausible view, I have sufficient reason to will that this not take place even if it costs me my life. Parfit makes a similar move when the reasons against the optimific principle are ones of partiality (Vol. 1, pp. 387–8). Given the scale of what is at stake in the universal acceptance of a principle, we have sufficient non-deontic reason to will the optimific principles even at great cost to us or those we love.

The most difficult case is one in which the features of an act that make it wrong are thought to give decisive reason not to will an optimific principle that requires us to act in that way. In Bridge, a runaway train will kill five people unless you cause me to fall in front of the train, resulting in my death (Vol. 1, pp. 390–1). According to the Wrong-Making Features Objection, the principle of saving five in Bridge is optimific, but there is decisive non-deontic reason not to save the five, and therefore not to will the optimific principle. Perhaps this reason lies in the fact that you would be harming one as a means to helping others. Parfit responds to this problem in three ways. He argues, first, that if the fact of harming one as a means to helping others gives decisive reason not to save the five in Bridge, there is non-deontic reason to will that others act accordingly (Vol. 1, pp. 391–2). On this assumption, the principle of saving five in Bridge would not be optimific and the objection would lapse. Parfit argues, second, that wrong-making features do not provide decisive reason to act in ways that violate the optimific principles (Vol. 1, pp. 394–5, 448–51). And he argues, third, that even if they did, they would not give decisive reason not to will these principles. We might have decisive non-deontic reason not to harm one as a means to helping others, but given the scale of what is at stake in the universal acceptance of a principle, we do not have decisive non-deontic reason to prefer that everyone act this way (Vol. 1, pp. 395–8).

There is a lot to say about these complicated moves, which have been the focus of much discussion. (Parfit responds to some of these discussions in an endnote, Vol. 1, pp. 476–9.) What strikes me about them is that they treat what we might call the Wrong-Making Principle — that when an act is wrong, the facts that make it wrong provide decisive non-deontic reasons — only as an objection to the convergence argument. Parfit does not consider the broader theoretical significance of this claim. My thought is that the Wrong-Making Principle threatens Kantian Contractualism in a more fundamental way. Parfit's formulation of this view asks what we have sufficient non-deontic reason to will. Why ignore deontic reasons? Because admitting them would undermine the view. In testing whether an act is wrong, we would have to determine, first, what there is deontic reason to will. That is, we would have to determine, first, which acts are wrong. Our test would be superfluous (Vol. 1, p. 287). There is a related problem if, corresponding to every deontic reason, there are decisive non-deontic reasons to act. Can we know what these reasons are before we know which acts are wrong? If not, we cannot apply the Kantian test. If so, we can know what there is decisive reason to do without applying this test, which is therefore redundant. (It would not help to extend the deontic beliefs restriction to preclude appeal to wrong-making facts, as well as deontic reasons, since that would rule out the application of the wide value-based objective view, which is essential to the plausibility of Kantian Contractualism). The upshot is that the value of Kantian Contractualism, its power to illuminate the content of morality, turns on the failure of the Wrong-Making Principle.

Now, Parfit does reject this principle, and he gives reasons for doing so (Vol. 1, pp. 394–5, 448–51). But these reasons are inconclusive, and there is a strong case for the Wrong-Making Principle, on grounds Parfit might himself accept. Suppose that, like Parfit, we hope to defend the rational authority of right and wrong. We believe, for instance, that an act's being wrong is a decisive reason against it (Vol. 1, p. 141). But that is not all. Consider this question: if right and wrong are of pivotal importance to practical reason, what should we make of an agent who is unable to conceive such facts, or who fails to consider them, or who is epistemically irrational in forming beliefs about them? Unless he is moved by the facts that make an act wrong without needing to form deontic beliefs, we should conclude that he is not ideally rational. If right and wrong have rational authority, a fully rational agent must recognize that an act is wrong when he knows the facts that make it wrong, and he must act on this belief, or he must act directly on the relevant facts. Either way, an agent who is not decisively moved by knowledge of wrong-making facts is less than ideally rational. It follows, through the connection between reasons and rationality, that the facts that make an action wrong provide decisive reasons against it. If right and wrong have rational authority, the Wrong-Making Principle holds.

This argument could be questioned in various ways. It depends, for instance, on a particular reading of rational authority, one that goes beyond the claim that deontic facts provide decisive reasons. But the argument is worth considering. And it raises a more general issue, about the starting point of Parfit's book. Why accept the wide value-based objective view? Why not begin with claims about non-deontic reason more congenial to morality, or less? Parfit argues at length against desire-based views of practical reason (Vol. 1, pp. 58–110). And his positive claims are not implausible. But why just these? Given Parfit's Non-Naturalism, there is no room for a metaphysical explanation: we cannot derive the content of practical reason from claims about its nature. That is fair enough. But if Parfit's assumptions about non-deontic reason seem plausible only because we accept the Wrong-Making Principle, or something like it, there is a deep instability in his approach.

It would be wrong to end on this sceptical note. *On What Matters* is one of the richest, most exciting contributions to moral philosophy in decades. Its ideas should energize others, even ones that Parfit himself rejects. For instance, I hope someone will take up his moving exploration of the Golden Rule, a moral principle that, despite its influence, is insufficiently discussed (Vol. 1, pp. 321–330). In doing so, they will be pursuing Parfit's hope for collaborative progress in moral thought. Parfit wants others to build on his work and to carry it further. His writing displays throughout a spirit of generosity and constructive engagement that is itself a moral ideal.

Department of Philosophy
University of Pittsburgh
Pittsburgh, PA 15260
USA
kis23@pitt.edu
KIERAN SETIYA

doi:10.1093/mind/fzs017
Advance Access publication 24 April 2012

Signals: Evolution, Learning, and Information, by Brian Skyrms. Oxford: Oxford University Press, 2010. Pp. vii + 195. H/b £30.00, $60.00; P/b £14.99, $27.00.

Suppose there are two agents. A *sender* can see the world but not act except to create signs of some kind that can be seen by a second agent. This *receiver* can act, but can only see signs sent by the sender. Actions by the receiver have consequences for both parties, and the two parties agree on which acts are good in each state of the world. By means of rational choice and common knowledge, agents such as these can maintain a sign system that seems to have at least rudimentary semantic properties.

This is David Lewis's model of conventional signalling, developed in his dissertation and 1969 book *Convention*, and intended as a reply to W. V. Quine's sceptical treatment of meaning. The model had limited influence on naturalistic philosophy, in part because Lewis presupposes rational agents who have thoughts with intentional properties. In a brief chapter in his 1996 book *Evolution of the Social Contract*, Brian Skyrms showed that evolution by natural selection, as well as rational choice, can give rise to signalling