DEREK PARFIT

Derek Antony Parfit

11 December 1942 – 2 January 2017

elected Fellow of the British Academy 1986

by

JONATHAN DANCY

Fellow of the Academy

Derek Parfit had acquired an international reputation by the time he was 35, and after the publication of his *Reasons and Persons* (1984) became one of the three or four most respected moral philosophers of his time. He eventually inherited Bernard Williams' position as the UK's leading moral philosopher. He was awarded the Rolf Schock Prize in 2014.

DEREK PARFIT

Biography

Derek Parfit was born in Chengdu, Szechwan, West China. His parents Jessie and Norman Parfit, both doctors, were medical missionaries; all four of his grandparents were missionaries too. His parents lost their faith when he was very young, but it was because of the war that in 1945 the family moved back to the UK. They went first to London and then to Oxford, where Parfit attended the Dragon School as a day boy; from there he won the top scholarship to Eton. At Eton he specialised in history, which was effectively the only alternative offered to Classics, which he detested. But it was at Eton that he first came across philosophy and this became the subject he most enjoyed; there was a philosophy club, to which practising philosophers were invited and at which Parfit read papers about free will and desert, God and the problem of evil. He also edited the Eton College *Chronicle*, and even left Eton a little early in order to take up a temporary post at the *New Yorker* during the summer of 1961, which he was offered out of the blue by the then editor, William Shawn; during that period he wrote a poem which was published in the *New Yorker* a year later.

From Eton, Parfit won a scholarship in history to Balliol College, Oxford, where he read History (1961–4). He also engaged in various journalistic activities, just as he had at Eton. But in his second term there he was already considering switching to Philosophy, Politics and Economics (PPE), which was the only way in which a non-classicist could study philosophy; what prevented him from doing so was that the PPE syllabus required one to study logic, which he hated and which made him feel very uncomfortable.

His intention on graduating was apparently to continue as an academic of some sort, since he took the examination for an All Souls Prize Fellowship in 1964 with history as his subject. He did not succeed, but the then Warden of All Souls, John Sparrow, wrote to him saying that he had been as close as a whisker to being elected and that he should consider trying again. He put that idea on the back burner, because he already had a Commonwealth (Harkness) Fellowship for the years 1964–6, which required him to travel in the US, and enabled him to take classes at Columbia University and Harvard University. (David Wiggins reports that Parfit somehow impressed the Harkness selection committee by talking about American conceptions of the future.) Parfit said later that his main motivation for taking up the offer of a Harkness Fellowship was that it gave him two years in which he could study whatever he wanted at American universities, and this would give him the time to decide whether to continue as an academic and whether to try to change his subject from history to philosophy. His experiences in the US led to his official conversion. The influence of Stuart Hampshire FBA (who himself had switched from history to philosophy, but during his undergraduate years) seems also to have been involved in this process.

A typescript of a paper of Parfit's on moral justification dated March 1967 has recently been unearthed, with a handwritten note '14 months after I changed my subject to philosophy'. This gives us a date of January 1966 as the moment of decision, with nearly half of the Harkness still to run.

So, on his return to the UK in January 1967, Parfit started as a graduate student at Balliol, reading for the BPhil in philosophy. He says in *Reasons and Persons* that he owes a great deal to his first teachers: Sir Peter Strawson, Sir Alfred Ayer, David Pears and Richard Hare (all FBA), who gave him tutorials during this period. But he sat the All Souls examination again at the beginning of the autumn term 1967, taking the philosophy papers this time, and was successful. (On election, he announced his subject as 'philosophy, in particular philosophy of mind and moral philosophy'.) Thus began his long association with the College, which he never left and which he loved so dearly.

On his election, he was urged by the Warden John Sparrow to withdraw from the BPhil, and so switched to do a DPhil entitled 'The philosophical concept of personal identity'. Alan Montefiore was his supervisor (1970–3)—an interesting choice; but in fact Parfit never got as far as submitting. This hardly mattered: he published his first article 'Personal identity' in January 1971, to considerable acclaim and from then on enjoyed an international reputation which made degrees an irrelevance. He did win the T. H. Green Prize in 1977 with an essay entitled 'Against Prudence'. But like other leading younger philosophers at Oxford (John McDowell FBA and Gareth Evans) he never got a graduate degree in Philosophy. At least the others had an undergraduate degree—in McDowell's case two of them.

One consequence of his peculiar path into philosophy was that Parfit was much less broadly based in the subject than most academic philosophers nowadays. He learnt up what he needed in order to pursue the topics that interested him. He did not need to concern himself with epistemology or logic or philosophy of language or philosophy of science, and he did not worry about such matters. There is a letter to the Warden of All Souls in 1971 in which he says that he wants to learn some philosophical logic. But I doubt that he went very far in that direction, for, as he confessed to me, or perhaps boasted, he had a sort of blindness where symbols are concerned; this was what had led to his decision not to switch to PPE at Balliol. So the general habit at that time of dressing up every argument in symbolic logic would have rendered much of what was written in certain fields impenetrable to him anyway.

Parfit's essay for the T. H. Green Prize was an interesting forerunner of what was to emerge later as *Reasons and Persons* (Oxford, 1984). It discussed cases where if each of us does what is best for ourselves, the result will be worse for everyone. It asked whether a rational person should be equally concerned with all the temporal parts of their future; and it discussed the implications for rationality of different views about the nature of personal identity.

All Souls suited Parfit perfectly, but a Prize Fellowship only lasted for seven years maximum, and after that, if one wanted to stay on, one had to apply for a Research Fellowship and then again for a Senior Research Fellowship. To be kept on one needed to be able to show significant achievement, and at each of these steps Parfit faced a challenge because he found it so hard to finish anything. In his application for a Research Fellowship (October 1973, towards the end of his Prize Fellowship) he promised three books. Material about personal identity was to form the heart of the first one: does personal identity have the importance we usually suppose, and is the concern for one's own future somehow specially rational? The second book was to be about rationality, with the aim of bringing the different kinds of reason for action within a unified account. The third book was to ask about societies what the other two books asked about individuals. But these promised books did not materialise. Already in 1975 Parfit had reported that he had got stuck and was considering writing a book entitled 'Future Generations', and thought he could only make progress by concentrating on rationality (which would include such notions as what is in one's own best interest and the incommensurability of values).

The non-appearance of the three promised books led to difficulties. In 1980, as his tenure of a Research Fellowship neared its end, Parfit applied for a Senior Research Fellowship to start in 1981, but there was considerable reluctance within All Souls to give him this position when he had not produced a book of any sort, let alone the three promised books. Eventually he was given a limited extension of his Research Fellowship until 1984, with encouragement, but also a warning that if no book (of suitable weight) appeared by then, he would be out. The pressure this put on him was enormous, but the result was *Reasons and Persons*.

A month after the book's appearance in 1984, to enormous acclaim, Parfit re-applied for and was granted a Senior Research Fellowship at All Souls, enabling him to continue to work there for the rest of his life.

In his application for a Senior Research Fellowship, Parfit stated that his philosophical interests lay in the most 'central and most debated subjects, such as the choice between conflicting views about what is in a person's own best interests, the many objections to consequentialism, different theories about liberty and justice, and the claimed incommensurability of different values'. He also suggests that work in epistemology and metaphysics, on the nature of time, free will, and on the kind of objectivity we can achieve could make the same sort of difference to our substantial beliefs about ourselves as could his own work on personal identity. He was nothing if not optimistic about the effect that good philosophy can have. But what actually happened was rather different.

In *Reasons and Persons* there is no sign of metaphysics (other than what might be called the metaphysics of persons) or of metaethics, and the only ethical theory that

makes a significant appearance is consequentialism. But by the end of the century Parfit's interests had moved to a focus on two main themes. The first theme is the question of whether three great philosophical understandings of ethics—consequentialism, contractualism and Kantianism—could not in the end be shown to be different expressions of the same position. (This fits his stated aim of considering 'the many objections to consequentialism'.) The second theme is an uncompromising realism about the moral and more generally the normative, as evidenced in the title he chose for his great three-volume work *On What Matters*.

So the next phase of Parfit's philosophical trajectory was one in which he gradually converted himself into a metaethicist. He had always been haunted by the thought that Hare's prescriptivism (and other forms of non-cognitivism) had the consequence that nothing really matters. There are things that we care about, but there is no possibility of mistakenly caring about things that don't matter, or not caring about things that do matter. Parfit wanted to fight this sort of metaethical nihilism with all the strength he could muster. *On What Matters* (Oxford) was the result; its first two volumes were published in 2011, with a third (posthumous) volume in 2017.

For the historian of ideas, and especially of Parfit's ideas, the path that led the author of *Reasons and Persons* to *On What Matters* is an interesting one. In the first book, as I said above, there is really no metaethics at all; it is all normative ethics, and the only ethical perspective considered is consequentialism, which grounds the right on the good. In the second book—which is a peculiar compilation with substantial contributions from other authors—it is pretty much all metaethics and the theory of reasons. The three names that now stand out are Kant, Sidgwick and Scanlon. Sidgwick had made a significant appearance in *Reasons and Persons*, but there are only two references to Scanlon in that book and Kant mainly appears in an extended example (in Part 3 on personal identity), which concerns the question whether Kant could have been born to parents other than his actual parents.

Part of the impetus which drove this change of focus was an attempt to respond to the various contributions to a collection I was putting together in the 1990s, consisting of responses to *Reasons and Persons*. With Parfit's enthusiastic help, I assembled an excellent list of contributors/critics and as their papers came in, Parfit set himself to work out his replies. And the replies grew and grew until it became clear that there was far more material here than could be accommodated in a single volume which included both the critical contributions and Parfit's responses. It got to the stage where three volumes were in sight, one on Parts 1 and 2, one on Part 3 and one on Part 4 of *Reasons and Persons*. Blackwell, who were publishing this volume, or rather these volumes, were very tolerant over several years as this process unfolded. But during that period Parfit began to recognise that he was also beginning to develop a framework within which all this might fit, and which would be far better presented as an

independent contribution than as a set of responses to anything. So he pulled out and the contributors' papers were published, very late and without his replies, in 1997, under the title *Reading Parfit*. Meanwhile he began to assemble the material that became *On What Matters*.

It was my privilege during that period to be involved in Parfit's construction of his metaethical position, and to find a friend. For several years we met pretty regularly to discuss philosophy—his and mine. He devoted as much care and attention to my efforts as to his; his enthusiasm was sometimes more than I could match, but the excitement of climbing the stairs to his room in the back Quad at All Souls remains with me.

Topics on which we had endless discussions during that period included Bernard Williams' paper on internal and external reasons, to which Parfit returned over and over again, and the arguments for metaethical naturalism; also the theory of motivation, his views on which emerged as 'Reasons and motivation' (*Proceedings of the Aristotelian Society, Supplementary Volumes*, 71 (1997), 99–146). In the latter there occurs one of the most striking examples of what later became a dominant motivation for Parfit, to show somehow that the proponents of positions that had been taken to be in opposition were 'really' in agreement. The example in this case was the contrast between psychologistic and non-psychologistic, or realist, conceptions of what it is to act for a reason. A psychologistic account will claim that 'motivating reason' is a term for whatever psychological states cause agents to act as they do, normally taking such states to consist of combinations of beliefs and desires; a realist conception will try to maintain that the reasons for which we act are such things as that this is too expensive, which is not a psychological state of anyone. Parfit maintained, very implausibly in my view, that these accounts are both right. There are two kinds of motivating reason: psychological states and matters of (supposed) fact. For normative purposes, it is more natural to think of reasons as matters of fact; for explanatory purposes, it is more natural to look for something capable of causing action, for which psychological states of the agent look like excellent candidates.

In addition to these private meetings, there were regular meetings of a small group consisting of John Broome, Roger Crisp, Brad Hooker, Philip Stratton-Lake, Parfit and myself (a sort of Oxford-Reading nexus). Later this group expanded with the arrival of Julian Savulescu, and it was then not long before it became so large as to be quite a different sort of occasion.

A personal note: after several years of regular train rides from Reading to Oxford and back, I began to find that I could not keep up. Other duties (of which Parfit had none) made it increasingly difficult to give his material the attention necessary if I was to contribute meaningfully to our discussions. And it didn't seem fair to take up so much of his time when I was getting far more than I gave; so I began to withdraw—

which I now greatly regret. Another disincentive, of a different sort, was the fact that Parfit revisited the same material over and over again in a continuous search for perfection, and the law of diminishing returns began to set in for me in a way that it never did for him. He was still reworking this material fifteen years later.

Peter Momtchiloff has shown me letters written during this period, in which Parfit, seeking a contract from Oxford University Press, reveals his current thinking about books he might write. In 1996 he offers five books. The last of these, to be titled *Truth, Evil and the Sublime*, is, he says, too remote to talk about. The first, to be titled *Practical Realism*, would defend a realist and irreducibly normative conception of practical reasons. The second would rework his conception of personal identity in response to the relevant papers in my *Reading Parfit*. The third, to be titled *On What Matters*, would discuss various questions about the goodness of outcomes and the rightness of acts, focusing on problems about distributive justice and future generations. The fourth, to be titled *Reasons, Persons, and Reality*, would be a reworking of various themes in *Reasons and Persons*. (Parfit never abandoned this dream of eventually publishing an improved version of *Reasons and Persons* which would enable the latter to go out of print.)

In 2003 (the year after he gave the Tanner Lectures at the University of California, Berkeley), Parfit's thinking about publication had changed, though no book had yet appeared. His main aim now was to finish two books. The second and the larger of these, to be called *Rediscovering Reasons*, would have one part on reasons, one on normative ethics (consequentialism, Kantianism and contractualism) and one on metaethics. The first and smaller one would be a shorter version of the second part of *Rediscovering Reasons*, to be published together with comments by Scanlon, Wood, Herman, O'Neill and Wolf. These comments (apart from those of O'Neill) all eventually appeared in volume 2 of *On What Matters*.

****

Parfit's life revolved around two completely engrossing interests. The first was philosophy and the second was photography, the latter counting as an avocation. Such time and attention (and money) that he did not devote to the former, he devoted to the latter. After a hard day doing philosophy he would regularly spend the small hours (11pm–1am) working on his photographs. He used to say that his real motivation for accepting visiting positions at US institutions was to earn money to pay for his photography—which required significant funds, not just because he himself owned expensive equipment (which needed to be upgraded all the time) but because he would only allow the most advanced technicians to print his photographs (at least until the arrival of Photoshop and other such programmes), and such people were charging

£1000 for a single print. Since it might take several prints before Parfit was completely satisfied, considerable resources were required.

Parfit's photographs, which were nearly all of St Petersburg or of Venice, were works of art, not just in the sense that they were intentional products of aesthetic value, but in the sense that he changed them, first by manual retouching (done by a professional) and later electronically, so as to make the shapes of the buildings more aesthetically pleasing. He made things look the way they should have been. He also took the people (and every sign of the twentieth century) out so that they didn't distract attention or detract from the buildings.

Since Parfit's death, the Heritage Museum at St Petersburg has expressed an interest in putting on an exhibition of his oeuvre—but only if an account could be given of the relationship between his philosophy and his photography. Unfortunately, such an account has proved very hard to give.

Frances Kamm has suggested to me that Parfit took up photography because he thought he lacked visual memory; photographs would serve as a surrogate for memory. His own view was that though he had no artistic talent he could recognise what was good when he saw it. So he found a way of producing good art, namely to take thousands of pictures and choose the best of them to put through the retouching process.

****

Returning to philosophy: Parfit's dedication to the subject went far beyond the norm, even among the most focused of us. He used to rise just before lunch, and then work without pause (other than for dinner) into the small hours of the morning (often, perhaps normally, ending with some time for his photographs). He found it very hard to go to sleep; he only achieved that at all by a carefully composed admixture of vodka and pills. But while he was awake he worked at a speed of which most are incapable. If you sent him a paper for comment, you would often receive those comments within a very few days, and the comments would often be as long as or longer than the paper you had sent. He achieved this partly by a simple ability to concentrate, and partly by a fierce determination to avoid as much as possible of anything that would interfere with the progress of philosophical enquiry. He always wore clothes of exactly the same type: a white shirt, dark trousers and a red tie; and he was extremely reluctant to give up time for such things as important family gatherings and unnecessary socialising. But commenting on the work of others did not count for him as dispensable interruption; it was part of what he was trying to do, not an interference in that. Philosophy was for Parfit a cooperative activity. And he assumed that others were as devoted to it as he was. All his books start with a very long list of people to

whom he was indebted. All of us do this, of course, to some degree; but in Parfit's case the debt was often very significant. Before the publication of his first book he sent copies of the (supposedly nearly) final draft to very many philosophers around the world—over a hundred, I believe—saying that though they had no duty of any sort to read this material and send him comments, still he would be very grateful for anything that they had to suggest. The list of those who did send him comments has fifty-six names. This was a completely new way of doing philosophy, one that only the most gifted of us could hope to practice; otherwise nobody would get anything done.

In addition to his own work, and to commenting on the work of colleagues, Parfit had a long series of eventually distinguished doctoral students, all of whom tell grateful stories of his support. One might have supposed that someone who held himself to such high standards of commitment to the subject would have been an impossibly demanding supervisor, but all reports show that he was deeply supportive and encouraging, in a way that enabled his students to become the best versions of themselves and involved no attempt to turn them into less talented versions of himself.

From mid-career onwards Parfit held a series of visiting positions at various senior departments of philosophy in the US: Harvard, New York University and Rutgers. Even if early on his main reason for doing this was that it enabled him to pay for the expenses of his photography, he had another reason later on, which was that it enabled him to give more to charity. And he was as generous with his time while abroad as he was in the UK, so long as it was for the sake of the subject.

Amidst all this concentration Parfit did find time to relax and listen to music, mainly classical (especially Bach and Wagner) but also some jazz—Lester Young and Billie Holiday were favourites.

In 2010 Parfit married his long-time companion, Janet Radcliffe Richards. He died suddenly and unexpectedly late at night in their house in London in the small hours of 2 January 2017.

Philosophical work

Parfit's first published piece was on the nature of a person, or personal identity. Published in the *Philosophical Review* in 1971, this paper is at the foundation of much that he wrote for a long time afterwards; it is wonderfully human, idiosyncratic and engaging—personal, even. On the first page he says:

> Some people … agree that our criteria of personal identity do not cover certain cases, but they believe that the nature of their own identity through time is, somehow, such as to guarantee that in these cases questions about their identity must have answers. … This first belief … has, I think, certain effects. It makes people assume that the

principle of self-interest is more rationally compelling than any moral principle. And it makes them more depressed by the thought of ageing and death.

Here we see already two central strands to Parfit's thought—strands which persisted to the end. The first is the link between discussions of personal identity and ethics, which is one of the main themes of Part 3 of his *Reasons and Persons*, and which I will lay out in more detail below. The second is the practical relevance and importance of philosophical enquiry. Parfit thought that correct philosophical thought could free us from an obsessive concern with ourselves and our weal and woe, since the distinction between self and others was not as stark as common sense takes it to be.

The paper continues with a series of bizarre (though now familiar) imaginary scenarios, such as one in which I can divide my mind into two streams, neither of which has access to the other for a period. At the end of that period, I can reunite those streams. How are we to describe this scenario? If the situation persisted, and I could not reunite my streams, we might say that there are now two people in one body; but if I can reunite them, we would be less tempted to say that. Parfit's point at this stage is that there does not need to be a fact of the matter. Of a similar case devised by David Wiggins, in which my brain is divided and each half transplanted into a new body, the best thing to say is that I *survive as* two different people, but not allow that I am either or both of these people (Parfit, 1971: 8). Identity is the limiting case in which there is only one person *as whom* I survive. But there is nothing so special about that, and in order not to be tricked by our language into supposing otherwise, we should give up the language of identity, and with it the belief that there are important questions which turn upon questions of identity. Such questions include the question of whether an action would be in my own (personal) interest.

Parfit's views about personal identity reappear in a much more advanced form as Part 3 of his first book, *Reasons and Persons*. This book is as a whole much more about persons than it is about reasons. And though there is plenty of talk about reasons, the focus, at least in Parts 1 and 2, is really on rationality. The final Part 4 is about persons and value; Parfit there reveals an extraordinary set of tensions within our conception of the value of a life. As far as I can see, these tensions—or contradictions—are not themselves derived directly either from his paper on personal identity or from the previous three parts of *Reasons and Persons*. In that sense, then, Part 4 is freestanding. The material in it was developed in and for a class which Parfit taught with James Griffin and Jonathan Glover for several years from 1968 onwards. Part 1 is not about reasons either—unless we take the view that the notion of a reason can be expressed in terms of rationality (which I don't). Parts 1 and 2 are about rationality. So reasons don't really enter into Parfit's picture until his second book, *On What Matters*.

The links between Parfit's first paper and the material in the book which appeared thirteen years later are various. The first is that he is concerned in Parts 1 and 2 of *Reasons and Persons* with two theories about how we should act. The first of these is S, the self-interest theory: S's central claim is that each person has only one rational aim, that outcomes go, for him, as well as possible. According to S, then, we each of us have our own rational aims, which can conflict. But there is also C (consequentialism), whose central claim is there is one ultimate moral aim: that outcomes be as good as possible. This claim gives each person the same aim, and it would seem that if one of us succeeds in this aim, others cannot be failing.

S acquires much greater force if we accept the intuitive distinction between oneself and others. But once we allow that the question 'will it be me who gets this benefit?' can be answered by saying 'to some extent yes, and to some extent no' the focus of S seems to be lost, or at least to waver. Parfit does not rely on this in Parts 1 and 2 of *Reasons and Persons*. There he argues directly against S, on an ordinary understanding of the self. He produces wild and wonderful cases where acting in accordance with S will definitely make one's life go worse. It can be the case that, though I am doing what S tells me to do, it is worse for me. And if we all do what S tells us to do, it can be worse for each of us. Actually, and unfortunately, the cases don't have to be weird and wonderful, as is shown by the standard example of overfishing. We will all (and each) do worse if we all do what S recommends than if we all acted in some other way.

One central aspect of Parfit's discussion of S and C is his use of the notion of the self-defeating and the distinction between two forms of defeat, direct and indirect. A theory is indirectly individually self-defeating if it is true that, if someone tries to achieve the aims that the theory gives him, those aims will be, on the whole, worse achieved. And a theory is directly individually self-defeating if, should someone successfully follow that theory, they will cause the aims the theory specifies for them to be less well achieved than if they had failed to follow the theory successfully. A theory is directly collectively self-defeating if, were each of us to follow the requirements of the theory, we would cause the aims it gives us to be less well achieved than if none of us had followed the theory. These various forms of self-defeat make available a new range of potential criticisms of well-worn theories of practical rationality.

Like so many others, I remember reading this material with a sense of doors opening before me. Suddenly moral philosophy became exciting in a way that it had not been before; it gained a dimension. At that time, it consisted almost entirely of metaethics. We were interested in the nature of moral thought and judgement and the possibility of moral facts. In *Reasons and Persons* Parfit paid almost no attention to such issues, though they are at the core of his second book *On What Matters*. An example of this focus (and perhaps of its limitations) is the notion of self-interest. On pages 3–4 Parfit gives several theories of self-interest, but none of these tell us what it

is for something to be in my interest—to benefit me, or be 'better for me'. All assume that this notion makes sense; in fact it is the subject of considerable dispute, but this is all metaethics (in the sense that it concerns the nature of a certain ethical concept, that of a benefit) and Parfit was not doing that. He did not examine his tools in that way.

Part 3 builds on Parts 1 and 2 in various ways. First, Parfit lays out his conception of personal identity through time as a matter of degree—to be more precise, as a matter of the degrees of psychological continuity and connectedness between the various temporal stages of the supposed self. Then he asks what difference this conception might make to our account of rationality as self-interest. The answer is that there is a discount rate. We should not be equally concerned with all the parts of our own future; the welfare of those parts of our later selves with which we are less well-connected counts for less. So it may not be irrational for a teenager to start smoking, knowing that doing so will make her later life much worse. The question is how much worse, and what is irrational is determined by whether the discount rate grounded in lesser degrees of connectedness is too steep, not steep enough or just right. 'Great imprudence is always sad, and often (as in the case of smoking) tragic. [But] we cannot claim that all such acts are irrational' (*Reasons and Persons*, p. 318). We cannot claim this because, on Parfit's account of personal identity, a teenager's adult self is *more like* a different person than we have standardly assumed. Rather, we should claim that such acts are not immoral. What is wrong with the teenager who starts smoking may not be that he is treating his later self irrationally. He is treating that self immorally, and that self is more like a different person than we normally suppose. Great imprudence is morally wrong even where it is not irrational.

But if it is morally wrong, this might affect the reasons we have for interfering. Autonomy does not include the right to harm your later self—or numerically distinct selves all equally well related to you—for no good reason. Further, since there will be no point at which I begin to exist, my existence being itself grounded in matters of degree, the morality of abortion may be affected (*Reasons and Persons*, p. 322). And someone in an incurable coma is not a person (because the degree of psychological continuity and connectedness is zero) and has no rights. So perhaps only the killing of persons is wrong. As for the morality of punishment, on Parfit's view a Statute of Limitations may be justified by the fact that identity is grounded in matters of degree.

These issues bring out the force in Parfit's original claim that decisions in the theory of persons can affect our moral views—in addition to our theory of rational action. Parfit also claimed that his belief that personal identity is grounded in matters of degree was liberating and consoling. It made him less concerned about his own future and his death, and more concerned about others (*Reasons and Persons*, p. 147). The passages in which he writes about these matters are some of the most moving, effective and humane to be found in modern philosophy.

Part 4 is a stand-alone tour de force. Parfit lays out two paradoxes in the theory of value—paradoxes to which nobody has yet succeeded in providing any resolution. He starts with what he calls the Non-Identity Problem, which is not itself a paradox.

The Non-Identity Problem concerns the question how one act or policy can be worse than another if it is not worse for anyone. Suppose we have two policies, A and B, and that different people will be born depending on which policy we adopt. We might want to compare the two policies and pronounce one worse than the other, but there is nobody for whom either policy will be worse or better. Parfit says that we need a theory, Theory X, which will answer this question. It might seem easy to provide such a theory: our answer might include the claim that if all the people who will be born if we adopt policy A will be happier than all the people who will born if we adopt policy B, policy A is better than policy B. This does not require us to say that adopting policy A is better for anybody; we can restrict ourselves to saying that the people born under policy A are better off than would be those who would be born under policy B. We don't need the notion of a benefit to say this; to benefit someone is to make something better, and that notion is inherently comparative (though what it compares is a matter of dispute). But the notion of being well off is not itself inherently comparative. So it is not clear that the Non-Identity Problem is on its own terribly hard.

But we also need to resolve two much harder issues. The first of these is the Repugnant Conclusion and the second is the Mere Addition Paradox.

We might agree that it is better if more happy people are living than if fewer equally happy people are living. In this sense, numbers can matter. But consider cases where, as the population grows, the quality of life falls—though it never falls so far as to make the lives of those living there not worth living. We might find ourselves having to choose between two outcomes, in one of which everyone is very happy, and in the other of which people are somewhat less happy but there are quite a lot more of them. Once we agree that the second world is better, there is considerable pressure to agree that a world with an enormously large population of people whose lives are just worth living is better than a world with a much smaller population of people whose lives are very well worth living. This better world is not better for anyone, of course, because the populations of these two worlds are quite distinct. But it is 'better in the sense relevant to choice': we should choose it if we can. This is the Repugnant Conclusion. A fairly grim world is better than a world of general bliss.

The Mere Addition Paradox asks whether lowering the average level of welfare is itself bad. Normally that would be so, because to lower the average level we would have to lower the level of some individuals. But if we enlarge our population by adding extra people whose lives are much worse than those of the original population, though still worth living, do we make things worse? This is not a non-identity issue.

We cannot hold that lowering the average quality of life is itself bad, if there is nobody whose quality of life is lowered, and if the means we take in order to lower the average quality amount to the creation of extra people whose lives are worth living. Nor does the greater inequality which we produce merely by adding more people make the new world worse than the one we started from; this is specially obvious if the new people don't know about the original population, for then there would be no issues about either social justice or jealousy. (I am not so sure about the social justice side of that.) The result is that it is always better to have an enormous population whose lives are only just worth living than a much smaller population whose lives are well worth living or even blissful. The numbers count.

We might hope that a path to the solution to (some of) these difficulties would be opened by increased subtlety in the theory of value. We might for instance decide that 'not worse than' does not imply 'at least as good as'. Parfit considers some such suggestions towards the end of Part 4. But he decides that they don't work.

Parfit's last publication 'Future people, the non-identity problem, and person-affecting principles' (2017) was submitted to the editors on 1 January 2017, only hours before his death, and is in fact a return to these issues, with a purported resolution. He is looking for a principle that does not encourage levelling down—rather, for one that discourages it. What emerges is a 'Wide Dual Person-Affecting Principle':

> One of two outcomes would be in one way better if this outcome would together benefit people more, and in another way better if this outcome would benefit each person more.

This principle supposedly implies that if N people exist at level 100, this is better than if 2N people exist at level 50; the benefits to each would be greater and the benefits to all would be the same. But the principle retains the idea that it is better to have 2N people at 100 than to have N people at 100.

In this way the levelling-down process that drives us to the Repugnant Conclusion is supposedly stopped. My own view, however, is that this resolution is not successful. Yet again, the paradox is stronger than we are.

An alternative suggestion, which appeals to me more, focuses on the relation between values and reasons. The Repugnant Conclusion is a result in the theory of value: a large enough world at a very low average level of well-being will always be better than a much smaller world with a much higher average level of well-being. But we should bear in mind the relevance of this to decisions about what we have reason to do. Parfit himself came to accept Scanlon's buck-passing conception of value, according to which the fact that something is of value is the same fact as the fact that we have reasons of certain sorts (to protect, promote … it). And his phrase (in *Reasons and Persons*) 'better in the sense relevant to choice' reveals a nascent tendency towards

what one might call a deontic conception of value, since 'relevant to choice' seems to mean something like 'relevant to what we ought to (decide to) do'. But if we reject such conceptions of value, we have room to suggest that though a repugnant world would be better than the present smaller world, that fact gives us no reason to prefer it, to work towards it if we can and so on. After all, the fact that a child, if I were to have it, would have a good quality of life is no reason to have that child—though it might perhaps intensify (or act as an enabler for) any reasons I do have to have that child. By contrast, the fact that, if I had a child, that child's life would not be worth living may be—probably would be—a reason not to have that child. But this need not determine our answer in the positive case.

Both of these remarks may seem peculiar because the reference of 'that child' in the question of how I have reason to act contains a reference to an object that does not yet exist. But we can sidestep that by saying that the fact that any child I have would have a life that is not worth living is a reason not to have a child—not to have any child. Analogous remarks on the positive case are less persuasive: the fact that any child I were to have would have a life worth living is not itself a reason for me to have a child—any child at all.

I now turn to Parfit's second major work, the massive three-volume *On What Matters* (volumes 1 and 2, 2011 and volume 3, 2017, comprising altogether 1,700 pages), which is dramatically different from his first, not in the stresses and strains that were involved in its production but in topic or rather in topics. The book derives in part from the Tanner Lectures which Parfit delivered at Berkeley in 2002. It has two main aims. The first is to suggest that three classic positions in normative ethics—Kantianism, rule consequentialism and contractualism—though they appear to be rivals, may in fact, with suitable adjustments, be reworkable as three facets of a possible single position. This suggestion is mainly treated in the last sections of Volume 1. The second aim is to establish a form of cognitive realism about the normative, and this is what the book's title refers to. Parfit had long been of the view that subjectivist conceptions of ethics commit their defenders to the claim that although some things may matter to us, indeed may matter very much to us, nothing really matters at all. All there is is what we care about, and though some carings may be more sensible than others or otherwise to be recommended, no carings are more correct than others. In this sense, nothing really matters. To say that we should care about some things is to express concern about our carings, not to enunciate some form of normative truth.

The early chapters of Volume 1 are devoted to the theory of reasons, something which, as I have said, was noticeably absent from *Reasons and Persons*. Parfit accepts Scanlon's so-called 'Buck-passing Account' of value, which sinks value into reasons; to be of value is to have features that give us reasons of certain sorts—reasons to care,

to conserve, to promote etc. And Parfit offers a raft of distinctions between various kinds of reasons. In particular, he introduces a topic to which he repeatedly returns, the distinction between objective and subjective theories of reasons. Subjectivists maintain that what we have reason to do is what would most achieve what we want; our reasons depend on our desires. Objectivists maintain that what we have most reason to do is what would be best. Parfit is an objectivist, but he is alarmed by, and repeatedly returns to, the views of people he respects who are of a subjectivist persuasion, most notably Bernard Williams.

But the first main aim of *On What Matters* was highlighted in the title Parfit originally gave to the working manuscript of his eventual trilogy, namely *Climbing the Mountain*. What Parfit does is to adjust all three theories (rule consequentialism, Kantianism and contractualism) so that their central claims can be expressed in a single formula TT:

> An act is wrong just when such acts are disallowed by some principle that is optimific, uniquely universally willable, and not reasonably rejectable. (p. 412)

The first criterion for wrongness here is consequentialist, the second Kantian and the third Scanlonian. Parfit's point, of course, is that these criteria need not be seen as conflicting. They may represent a difference of focus, but there is no difficulty in supposing that a principle might satisfy all three of them at once.

My immediate worry, when I first read this material, was that Parfit had carefully omitted to include the intuitionist tradition in his list of positions to be accommodated. Ross's (1930) theory of *prima facie* duties, a theory which seems to me to merit inclusion on any list of classic positions in normative ethics, gives a quite different account of which acts are wrong. For Ross, an act is right if it is one's most pressing *prima facie* duty. Showing that one could, at a pinch, squeeze all the non-Rossian theories—by which I mean theories that work in terms of absolute duties rather than in terms of Ross's *prima facie* duties—into one combined principle didn't fully cut the ice for me. For Ross's *prima facie* principles don't disallow or allow, require or forbid anything; there is no such thing as 'being *prima facie* disallowed' or 'being *prima facie* required'. And Ross's principles do not emerge as 'not reasonably rejectable' in anything approaching Scanlon's sense of that phrase.

What is more, even if we allow an evaluative shadow of Ross's apparently non-evaluative *prima facie* principles, it seems that there is no such thing as being *prima facie* good or *prima facie* bad—though Ross does suggest otherwise occasionally. An act can be *prima facie* right or *prima facie* wrong, and it can be good in a respect and bad in another respect, but being good in a respect is not a form of *prima facie* goodness; it is real, actual goodness—unlike the rightness in *prima facie* rightness, which is not a form of rightness at all.

But in the penultimate chapter of the entire book, chapter 57, Parfit does address Ross's view directly. He argues there that there is a possible position which he calls Act-including Act Consequentialism, which holds that

> C: it would often be intrinsically bad to treat people in certain ways, such as deceiving or coercing them, or breaking promises we have made to them.

Technically this is not a form of consequentialism, since it does not mention the wrong, and does not hold that all wrong acts are wrong because of the disvalue of their consequences. But it is still a view which holds that some of our *prima facie* duties are grounded in the goodness or badness of *something*, namely the goodness or badness of the relevant action; and for many that would be enough for them to think of it as a form of consequentialism. We would have to avoid the temptation to say that these acts are good because they are right or bad because they are wrong. If this is a form of consequentialism at all, it has to go the other way around: they are right because they are good and wrong because they are bad. The question then is what sort of value we are dealing with here. Parfit says that it is a form of intrinsic value. But are we to think of this value as moral value or just as ordinary goodness? Standard consequentialism understands the right and the wrong in terms of non-moral value: the value of happiness, friendship, achievement and so on. Act-including Act Consequentialism seems to understand the right and the wrong in terms of a new form of value, the intrinsic (not consequential) moral value of an action. Of course, we are used to moral value if it is a value that belongs to agents: some agents are morally good and others are morally bad. But this new sort of moral value is to be a value that belongs to actions, not to agents. And certainly it is a good thing that right actions be done. But this sort of goodness is not moral goodness; it is the ordinary sort of goodness that we might express by saying that this or that behaviour was welcome. So the question remains of whether we can make sense of this new sort of moral goodness and badness that belongs directly to actions and is grounded in such features as deceit and coercion and charity.

Matters are complicated by the word 'often' in Parfit's account of Act-including Act Consequentialism. A principle is not something that applies 'often'. A principle applies universally or not at all. There are such things as contributory principles; but they specify a feature that makes an invariant contribution, one that is always there and will sometimes win the day. There is no 'often' about it.

One way forward would be to distinguish between intentional deceit and unintentional deceit. Intentional deceit looks as if it is a feature of the agent and any disvalue it has will belong to the agent as well as to the action. So what about unintentional deceit? We might understand this as unintentionally causing someone to have a false belief. This is no doubt often (but not always) unfortunate and unwelcome, but

it doesn't seem to be morally bad as such; and if it is bad, it will be bad because of some consequence—and we are not supposed to be appealing to the badness of consequences here.

It is hard to see one's way through this thicket, but my general conclusion is that Parfit did not manage to accommodate Ross-style intuitionism. His mountain has three sides when it needs really to have four. More generally, anyone familiar with Ross's *The Right and the Good* will be struck by the fact that none of the potential moral principles that Parfit considers are *prima facie* principles in Ross's sense. Parfit deals with Ross by proposing a further non-*prima facie* duty which is not to do with the value of consequences. (Though the word 'often' in C does again give one pause.) Though Ewing (1959, p. 126) called Ross's conception of a *prima facie* duty 'one of the most important discoveries of the century in moral philosophy' I don't think that Parfit ever took that 'discovery' on board.

I now turn to moral metaphysics, which occupies very much more space, as is appropriate in a work entitled *On What Matters*. Here the question is not which things matter and how much, but whether anything matters at all.

Parfit's treatment of the relevant metaethical material in volumes 1 and 2 differs from what we find in volume 3, which appeared six years later, but the changes are not enormously significant. His general picture of the normative, which he used to call Non-Metaphysical Non-Naturalism, is now called Non-Realist Cognitivism. I can't myself see that the change of name is any improvement; indeed, it might be a step in the wrong direction. The term 'Non-Metaphysical' showed, perhaps correctly, that there was to be no challengeable Metaphysics; 'Non-Realist' merely avoids a particular metaphysical picture. And the term 'Cognitivism' misleadingly directs our attention to issues which concern the nature of an appropriate response and away from the nature of that to which we are responding. Perhaps all that is meant is that the notions of true and false, of getting it right and getting it wrong, are appropriate in ethics and in normative enquiries more generally. Parfit's aim here is to allow that the distinction between true and false applies to moral claims and utterances just as much as to claims and utterances in natural science and indeed in mathematics. But this can only be because the applicability of the true/false distinction is not a metaphysical one and does not depend on metaphysics. We tend to think that a true remark or thought must be one that fits the facts, and that it is the part of the naturalist to insist that the only facts are natural facts. Parfit is using a supposedly non-substantial sense of 'fact' in which there are moral facts as well as natural and mathematical facts. But these moral facts do not obtain because of the existence of anything like a moral state of affairs. There is truth here but it is not a form of correspondence to some sort of independent moral reality, because there is no such reality.

What is it then? Parfit did not seem to be interested in this question. It was enough for him to fend off the opposition (in the form of Alan Gibbard, Simon Blackburn FBA and others). But the rest of us might be looking for more than this. The work of David Wiggins FBA and John McDowell FBA in the 1970s springs to mind, especially Wiggins' 'Towards a sensible subjectivism' (not published until 1987). The most that Parfit gives us is the term 'Cognitivism', and this is a term which he does not attempt to unpack. But one would have thought that a moral belief is a form of practical commitment, and something of that sort is not happily characterised simply as cognitive, even though the main argument for what we used to call 'realism' was that there is a constant fear of getting it wrong (that is, of making moral mistakes) despite one's own best efforts. Of course, the question then is what sort of mistake we make when we get it wrong.

In fact, Parfit's metaphysical position is very close to John Skorupski's notion of irrealism: 'Irrealism about the normative says that normative knowledge of fundamental normative propositions *rests on no receptive awareness*. The only capacity it requires is the non-receptive cognitive capacity of rationality, a capacity which involves spontaneity and regulation by the universality of reasons, not receptivity' (Skorupski, 1999, 456). This idea of non-receptive awareness of a truth is one that Parfit might have found fruitful. But we would still be left looking for a positive account of what 'non-receptive awareness' amounts to, since knowing what it is not does not tell us what it is. One final suggestion is that 'non-receptive' means, or at least entails '*a priori*'; on this view, our moral knowledge would emerge as *a priori* knowledge of particular normative truths, consisting mainly in truths about reasons. But this is only a gesture towards an adequate epistemology.

References

Dancy, J. (ed.) (1997) *Reading Parfit* (Oxford).

Ewing, A. C. (1959) *Second Thoughts in Moral Philosophy* (London).

Parfit, D. (1971) 'Personal identity', *The Philosophical Review*, 80, 3–27.

Parfit, D. (1984) *Reasons and Persons* (Oxford).

Parfit, D. (1997) 'Reasons and motivation', *Proceedings of the Aristotelian Society, Supplementary Volumes*, 71, 99–130.

Parfit, D. (2011) *On What Matters*, vols 1 and 2 (Oxford).

Parfit, D. (2017) 'Future people, the non-identity problem, and person-affecting principles', *Philosophy & Public Affairs*, 45, 118–57.

Parfit, D. (2017) *On What Matters*, vol. 3 (Oxford).

Ross, W. D. (1930) *The Right and the Good* (Oxford).

Skorupski, J. (1999) 'Irrealist cognitivism', *Ratio*, 12, 436–59.

Wiggins, D. (1987), 'Towards a sensible subjectivism', in his *Needs, Values, Truth: Essays in the Philosophy of Value* (Oxford), pp. 185–214.

Acknowledgements
I am very grateful to John Broome, David Wiggins, Jeff McMahan and Janet Radcliffe Richards for their help. I also wish to thank the Warden of All Souls, Sir John Vickers, for giving me access to the College's file on Parfit, from which most of the details of Parfit's time at All Souls are taken.

Note on the author: Jonathan Dancy is Professor of Philosophy at the University of Texas at Austin. He was elected a Fellow of the British Academy in 2016.