DOI: 10.1111/j.1467-9205.2012.01470.x
Philosophical Investigations ••:•• •• 2012
ISSN 0190-0536

CRITICAL NOTICE

Climbing Which Mountain? A Critical Study of Derek Parfit, *On What Matters* (OUP 2011)

T. Chappell, *The Open University*

What does ethical sensibility need from moral philosophy? What should it want or expect, if its expectations are realistic? What method(s) and style(s) are appropriate for moral philosophy, and how do these best serve the ends of moral philosophy? What, above all and most basically, is moral philosophy *for*?

These are fundamental questions for professional moral philosophers, and they impose themselves relentlessly on us the moment we step outside the seminar room. Inside the seminar room, playing along under current moral philosophy's normal conditions, it is remarkably easy to lose sight of them. Moral philosophers mostly seem to be in the business of giving answers and expend endless energy and ingenuity on comparing and criticising each other's answers. But to state a point that is no less crashingly obvious in moral philosophy than anywhere else, to know what the right answer is, we need to know the question. As we can also hardly avoid noticing, there may be more than one right question.

Which question we are answering – which mountain we are climbing, to use Parfit's own lead metaphor – not only determines what counts as a good answer but also determines whether or not the familiar "main schools of moral theory" should be taken, as they routinely are, to be competitors. Suppose, for example, that the utilitarian is asking "What is it best for me to do?," the Kantian "Which if any currently possible and salient actions have I conclusive reason *not* to do?," the contractarian "How should we justify our actions to each other?," and the virtue ethicist "How should I live?." So far forth in this possible scenario, Kantians, utilitarians, contractarians and virtue ethicists are not even on the same mountain – not even trying to answer the same question.

My point is not that this scenario is actual; some moral philosophers may see things this way, others clearly do not. My point is rather that the

possibility is real and insufficiently noticed.[1] Its particular interest here is that it provides a way of running a completely different conciliatory project between Kantianism, utilitarianism and contractarianism from the project that (*inter alia*) concerns Parfit in his large, wide-ranging, important, splendidly clear, deeply learned, good-humoured and entertaining book *On What Matters* (OWM).

However, just to get this far, we have to make some crucial assumptions that it might be worth our while to pause and question. For instance, we have known since Socrates' time that it is not inevitable to assume that moral or other philosophy, whatever questions it may *raise*, has to be in the business of answering them. (Simply answering them, or mainly answering them, or answering them at all.) Wittgenstein famously said that "Philosophy is not a body of doctrine, but an activity" (*Tractatus* 4.112), and that "If one tried to advance *theses* in philosophy, it would never be possible to debate them, because everyone would agree to them" (*Philosophical Investigations* I, 128). Again, and perhaps less familiarly, Simone Weil wrote this:

> The distinctive method of philosophy consists in getting a clear conception of insoluble problems in their insolubility, then in contemplating those problems without anything else; fixedly, tirelessly, for years, without the least hope, in a state of waiting.
>
> By this criterion, there are few philosophers.

Weil's idea that philosophy – perhaps particularly moral philosophy – might be essentially aporetic is one insufficiently considered idea; Weil's idea that it might be essentially contemplative is another. It is not hard to see other possibilities less extreme than these (and so, maybe, more plausible – like Parfit himself (OWM I: xliv), Weil loves extremes). One is the idea that we cannot hope to answer moral questions well or wisely unless we have gained some sense of their genuine and sometimes almost overwhelming difficulty. Another is the idea that how one answers a moral question may not be the only, or the most, important or interesting part of the process of moral reflection. Other, perhaps greater, interest and value may lie in the reflective process itself rather than its outcome, and in particular, in what one's conduct of that process reveals or creates in one's character.

I think there is something in both these latter ideas. To develop the second, the idea of moral philosophy as contemplation, consider the phenomenology of a part of life that must be familiar to anyone in academia (although of course not only to them): the activity of study.

1. Parfit notices it: "those who use 'wrong' or 'ought' in different senses may *not* be disagreeing," OWM I: 171.

There is a particular kind of calm, attentive, steady and focused collectedness that we achieve when we are studying well. This is the state of mind in which we understand things most clearly and learn things most easily; it is the kind of tranquil lucidity in which real understanding, genuine intellectual progress, most readily comes to us. It seems to be this mode of being and experience, and in particular, its essential freedom from selfishness, its other-directedness, that Simone Weil and Iris Murdoch have in mind when they speak of "attention." I think some similar mode of experience, and the natural human appetite for it (an appetite which, to repeat, is obviously not restricted to academics and students), must be what Bernard Lonergan means when he talks about "insight."[2] Undoubtedly, something related is what people get out of the form of "meditation" best known by that name in our society. Such outwardly directed lucidity is, as Weil notes, both rare and elusive; just how elusive it can be is perhaps suggested by Bertrand Russell's wonderful remark that true philosophy is something that a really good philosopher does about once in 6 months, a bad philosopher never.[3] The idea that this attentiveness and reflectiveness is one of the most valuable states that human beings can ever be in, and that this inner stillness is the essential precondition of any genuinely worthwhile action, is central both to the contemplative traditions of the great religions and also to the thought of Plato and Aristotle. It is also, I believe, entirely correct.

If what moral philosophy had to offer us was (*inter alia*, of course) ways of reaching, exploring and enhancing this serenely happy mode of being and experience, that would, to my mind, make moral philosophy an entirely comprehensible activity. Moral philosophy would then provide us

2. "Deep within us all, emergent when the noise of other appetites is stilled, there is a drive to know, to understand, to see why, to discover the reason, to find the cause, to explain. Just what is wanted, has many names. In what precisely it consists, is a matter of dispute. But the fact of inquiry is beyond all doubt. It can absorb a man. It can keep him for hours, day after day, year after year, in the narrow prison of his study or his laboratory. It can send him on dangerous voyages of exploration. It can withdraw him from other interests, other pursuits, other pleasures, other achievements. It can fill his waking thoughts, hide from him the world of ordinary affairs, invade the very fabric of his dreams. It can demand endless sacrifices that are made without regret though there is only the hope, never a certain promise, of success. What better symbol could find for this obscure, exigent, imperious drive than a man, naked, running excitedly crying, 'I've got it'?" [Bernard Lonergan, *Insight* (New York: Longman's, 1957), p. 4].

3. "[T]he subject-matter that you are supposed to be thinking of is so exceedingly difficult and elusive that any person who has ever tried to think about it knows you do not think about it except perhaps once in six months for half a minute. The rest of the time you think about the symbols, because they are tangible . . . The really good philosopher is the one who does once in six months manage to think about it. Bad philosophers never do" [Bertrand Russell, *The Philosophy of Logical Atomism*, Lecture 1, at p. 185 in R. C.Marsh, ed., *Logic and Knowledge* (London: Allen and Unwin, 1956); the remark is specifically about philosophical logic, but perhaps the extension is permissible].

with help in attaining something like what, in an earlier publication in this journal,[4] I called an "ethical outlook": an overall way of understanding life, and the meaning of life, that could reasonably be achieved and valued – by any reflective person, not just by academics or intellectuals – both for its own sake, and also as a way to help us to engage with the human world (and the other worlds) around us. Perhaps some philosophers worth classifying as "moral" do offer or attempt to offer something like this, even today (I doubt that any first-rank "moral" philosopher of the ancient world was ever centrally concerned to offer anything else). Perhaps Parfit himself gives us more unofficial help of this sort, in OWM, than we might expect. But what dominates the landscape of moral philosophy today is something rather different, and something much less comprehensible – to me at least, given my scepticism about the idea that systematic moral theory could provide us, all on its own, with anything like a credible ethical outlook. This is the search for a "theory of the right" in isolation from any of the background in personal reflection and personal commitment that alone could give any such theory real intelligibility. At least at first sight, this rather impersonal and rootless quest for a theory of the right is also what dominates a large proportion of the two fat volumes of OWM. ("At first sight": in fact Parfit is doing something less like assertion and more like exploration of possibilities. This needs to be borne in mind in what follows.)

Derek Parfit has a clear and straightforward view of what moral philosophy is for, and how it should proceed. As his title tells us in effect, the point of moral philosophy is to determine and define, at least in large part, "what matters," and hence, at least in that part, what our reasons are. It is a central concern of his, especially in Part Six, to argue that we *really do* have such reasons; irrealists are wrong to think that our reasons are in some way ultimately illusory. His campaign of argument against irrealism is a *tour de force*. Amid the forests of depressingly oblique, hyper-technical and often deeply boring discussions that have grown up in meta-ethics, Parfit has a refreshing ability to cut to the chase. I take the testiness of most irrealists' responses, to date, to Parfit's defence of moral realism to be a backhanded compliment. But for reasons of space, I would not engage closely with these arguments here.

As for the method of moral philosophy: What Parts Two to Five of OWM show us by example is that (at any rate) a method of moral philosophy is to accumulate precise arguments on particular points to look for convergence between theories. The omega point of such convergence

4. Timothy Chappell, "Ethics beyond moral theory," *Philosophical Investigations* **32.3** (July 2009), 206–243.

would be to find the single complete and objectively correct theory of the right, or to get as close as we can to that single theory. I have said elsewhere why I suspect that such a theory is impossible, is not worth aiming at even if possible, and would not be much use to us even if achieved.[5] However, Parfit sees these difficulties too; he denies that he is trying to go that far and also denies, as noted above, that he is actually *asserting* the converged view that his arguments lead him to, rather than, perhaps, just mooting it as a possibility. (And why bother doing that? Mark Schroeder[6] has suggested what seems to me the right answer: Parfit wants to show by example that moral progress is possible. If we can show that a wide range of serious and intelligent moral theorists have reason to see themselves as tending to converge on something like the triple theory, that is a reason for optimism about moral progress.)

Parfit and I also (he tells me) both agree with some related remarks that Thomas Baldwin makes about Parfit's "master" (OWM I: xxxiii) Henry Sidgwick:

> [S]ince human affairs are not systematic it is a mistake to think that the ordinary, common-sense morality which regulates and inspires human affairs needs to be "systematised" (and then "corrected"). Instead, in developing a critical appraisal of it, the place from which to start is with the thought that our common-sense moral judgements concerning our duties, responsibilities, and virtues are embedded in a largely implicit understanding of our culture and institutions: the family, place of work, school, shop, hospital, neighbourhood, state, etc. So a critical understanding of these judgements needs to be informed by an explicit understanding of those institutions and their place in human life, together with an awareness of the possibility of organising human affairs in other ways [Thomas Baldwin, in Philip Stratton-Lake, ed., *Ethical Intuitionism: re-evaluations* (OUP 2002), at p. 111]

Parfit's book does not give any focused consideration to such unsystematic possibilities for moral philosophy as Baldwin here describes, but that is not because Parfit rejects them. On the contrary, he explicitly leaves the door open to such possibilities when he writes, for example, that

> For some moral theory to succeed, it must have plausible implications. The Triple Theory has many such implications. But after we have worked out what this theory implies, and we have carefully considered all of the relevant facts and arguments, this theory might conflict with our intuitive beliefs about the wrongness of certain acts. If there are many such conflicts, or these intuitive beliefs are very strong, we could then justifiably reject this theory (OWM I: 415).

5. See reference in last note.

6. In his Notre Dame Philosophical Reviews notice of OWM, August 1 2011, http://ndpr.nd.edu/news/25393-on-what-matters-volumes-1-and-2/.

– where rejecting the triple theory does not necessarily mean rejecting it in favour of some other theory; it might, as Parfit acknowledges, just mean rejecting all systematic theories.

Despite these points, it is easy for people to read OWM – people do read it this way; Parfit has corrected me for so reading it myself – as if the only alternatives it really takes seriously are Parfit's own objective, which is taken to be a single true systematic theory of the right, on the one hand, and on the other hand, the threatening chaos of moral nihilism. Apart from accuracy as Parfit exegesis, such a view will not do because the opposition "systematic theory or irrealism" does not withstand scrutiny. Why could not a moral irrealist be a systematiser? Why, indeed, would not an irrealist have more need of system than a realist, inasmuch as he has nothing else except something like a criterion of coherence to hold things together? Think here of the way Simon Blackburn uses such a criterion to hold together systems of attitudes (see, e.g. *Spreading the Word*; Oxford: Clarendon, 1984, Chapter 5).[7]

With this (mis)reading of Parfit, compare a notorious argument from *The Methods of Ethics* (London: Allen and Unwin, 7th edition 1907, p. 406):

> If, however, this view [hedonism] be rejected, it remains to consider whether we can frame any other coherent account of Ultimate Good. If we are not to systematise human activities by taking Universal Happiness as their common end, on what other principles are we to systematise them?

If we reject hedonism, then we cannot "systematise human activities"; so, we must not reject hedonism. That we might prefer not to "systematise human activities" at all is an option that Sidgwick does not even mention here, although it can hardly have failed to occur to him. "To suppose that, if [the moral theorists'] formulations are rejected, we are left with nothing is to take a strange view of what in social and personal life counts as something" [Bernard Williams, *Ethics and the Limits of Philosophy*; (London: Harmondsworth, 1985), p. 200]. The option does, however, occur to Parfit, and like me, he criticises Sidgwick for missing it:

> Though Sidgwick once wrote "I will not stir a finger to compress the world into a system", he later did that. Sidgwick writes: "If we are not

7. Think too, to deploy the comparison that most usually haunts meta-ethics, of the ease with which at least minimally realistic science can accommodate mess and disorder: "What is scientific method? Is it the experimental method? The question is wrongly posed. Why should there be *the* method of science? There is not just one way to build a house, or even to grow tomatoes. We should not expect something as motley as the growth of knowledge to be strapped to one methodology" [Ian Hacking, *Representing and Intervening* (Cambridge UP 1973), p. 152].

> to systematise human activities by taking Universal Happiness as their common end, on what other principles are we to systematise them?" (*Methods of Ethics* 406). He should not have assumed that we *are* to systematise these activities, and that we should therefore be Hedonists (OWM I: 453, note on xl).

Still, in the forests of commentary that have already appeared, Parts Two to Five[8] of OWM have been widely understood as concerned with a simple central thesis: that the project of moral theory can be vindicated because properly understood the three most important systematic moral theories – contractualism, Kantianism and consequentialism – converge upon the single "grand unified" moral theory which Parfit allegedly advocates. In fact, as I say, he does not advocate it; it would be closer to the truth to say that he puts the triple theory up for discussion. Parfit himself is aware of, and exploits, the possibility that I mentioned at the outset, that moral philosophy might find other things worth doing with questions than simply and assertorically answering them.

Why, incidentally, is it a *triple* theory rather than a *quadruple* theory – why has virtue ethics dropped out as something to be made part of the convergence? The answer is that virtue ethics is either systematic or unsystematic. Unsystematic virtue ethics is not part of Parfit's discussion because, while as we have seen he is open to unsystematic possibilities, they are not his main concern in OWM. As for systematic virtue ethics, Parfit writes this (OWM I: 375):

> *Motive Consequentialists* similarly claim that, though the best motives are the ones whose being had by everyone would make things go best, the best or right acts are not the acts that would make things go best, but the acts that would be done by people with the best motives. These theories overlap with those systematic forms of *virtue ethics* which appeal to the character-traits and other dispositions that best promote human flourishing or well-being. There could be many other forms of Indirect Consequentialism.

Parfit thinks that systematic virtue ethics requires no separate discussion because, *taken as a theory of the right*, it is just a version of motive consequentialism. That seems to me a conclusive reason not to take virtue ethics as a theory of the right. For the equation strikes me, as I think it will strike most virtue ethicists, as false. Of course, there are consequentialist virtue ethicists, but for most virtue ethicists – Foot, MacIntyre, Anscombe, Williams, Hursthouse, Swanton and me – their view is a root-and-branch rejection of consequentialism, or it is nothing.

8. Including part 4, which is not by Parfit but by four distinguished commentators – Susan Wolf, Barbara Herman, Tim Scanlon, Allen Wood. Part Five is Parfit's responses to them.

One way of motivating this rejection is to point to the consequentialist's assumption that all reasons arise from the future.[9] That assumption seems groundless. There seems no obstacle to reasons equally arising from the past (think of promise keeping, gratitude, resentment and punishment), or from the present (think of welcoming a visitor or proposing an impromptu toast), or arising with no particular time index at all (think of the sorts of activities that Aristotle calls *praxeis*, e.g. playing music or, once more, study). One difference between the consequentialist and the virtue ethicist can be that the latter gives due recognition to this variety.[10] But it will hardly be the only difference if, as I have argued (and as Parfit tells me he agrees), virtue ethics is not well understood as a systematic theory of anything, so a *fortiori* not as a systematic theory of the right.

What, anyway, is the main argument for the "triple theory" of the right? (Note again – "*the* main argument," not "*Parfit's* main argument." Parfit is entertaining this theory, not asserting it.) Jacob Ross summarises the argument roughly as follows:[11]

> R1. *The Formula of Universally Willable Principles*: An act is wrong unless such acts are permitted by some principle whose universal acceptance everyone could rationally will.
> R2. There are some principles whose universal acceptance would make things go best.
> R3. Everyone could rationally will that everyone accept these principles.
> R4. At least in most cases, these are the only principles whose universal acceptance everyone could rationally will.[12]
> R5. So these are the principles that everyone ought to follow.

---

9. I develop this argument further in "Intuition, system, and the 'paradox' of deontology," in L. Jost and J. Wuerth, ed., *Perfecting Virtue*, Cambridge UP 2010. Notice the difference between saying that all reasons arise from the future – which, as my main text argues, is false – and saying that all reasons are future-directed – which must be true, at least in the sense that by its very structure deliberation is about the question "What *shall* I do?". I speculate that one of consequentialism's spurious attractions is its elision of "arising from the future" with "future-directed". (Thanks to Adrian Moore for discussion.)

10. Parfit addresses something like this point at OWM I: 373–4: "The word 'consequentialist' is in one way misleading, as is talk of the goodness of *outcomes* and the acts that *make* things go best . . . the goodness of some outcomes might depend in part on facts about the past. It might be better, for example, if benefits went to people who had earlier been worse off . . . And some acts, intentions, and motives may be in themselves good or bad, whatever their effects . . . When we ask whether it would be best if something happened, or if someone acted in some way, we are asking what, from an impartial point of view, everyone would have most reason to want, or to hope." If this means that Parfit himself rejects the thesis that all reasons arise from the future, then so far forth his own view is not consequentialist – though it does have other consequentialist characteristics, marked here by the words "from an impartial point of view" and "most".

11. Jacob Ross, "Should Kantians be consequentialists?," pp. 126–135 of the *Ratio* Special Edition on Derek Parfit's *On What Matters*, edited by John Cottingham and Jussi Suikkanen, April 2009. Rarely has so much been published about a book that has not yet been published.

12. In the light of comments that Parfit has made to me in correspondence, I have altered what Jacob Ross originally says by adding "at least in most cases" to R4.

And these principles are the substance of what Parfit (OWM I, S49) calls the

> *Triple Theory*: An act is wrong just when such acts are disallowed by the principles that are optimific, uniquely universally willable, and not reasonably rejectable.

R2's last three words imply that there is something determinate called "things going best": there is some one specifiable way the world *in toto* can be which counts as its best state. But this, it seems to me, is quite untrue. The point is, of course, a very general one, and it affects all sorts of moral theories, not just the triple theory (and not even just consequentialist theories). Still, it may be worth dwelling on a little.[13]

There is something bizarre about the very idea of *any* assessment of how well the world *in toto* is going. When I ask you "How are things?," this is not a vernacular ellipsis of "What is your evaluation of the current overall state of the universe?." As a matter of fact, we do not perform such evaluations; as a matter of principle, I doubt we know how to, or need to know how.

There is something particularly bizarre about the very idea of an assessment of some possible state of the world *in toto* as the *best* possible state of the world *in toto*. What would that mean? I do not know what a (finite) description of the world in its best possible state would sound like. I have even less idea why, for any such finite description, it would not always be possible to add some more details that make it *better*. There is no largest integer; why should there be a best total state of affairs? (Remember Gaunilo's island.)[14]

A standard response to this point is to retreat from the formulation "things going best" to the formulation "things going at least as well as in any other scenario we can imagine." But here again, the last three words are fatal. What have the limits of our imaginations to do with anything? (Whose imaginations, anyway?) And, anyway, this formulation fixes

---

13. For more about this argument see my "Option Ranges," *Journal of Applied Philosophy* **18** (2001), 107–118. See also Eric Wiland, "Monkeys, typewriters, and objective consequentialism," *Ratio* **18** (2005), 352–360, and James Lenman, "Consequentialism and cluelessness," *Philosophy and Public Affairs* **29** (2000), 342–370.

14. Parfit tells me (in correspondence) that his view is that "When we are comparing several possible outcomes, we don't have to be considering all other imaginable outcomes. I believe it would have been much better if the First and Second World Wars had been avoided, or if Aids had not killed millions of people. If you [TC] have some such beliefs, we don't disagree." I disagree; I think we do disagree. To think that there is something (or some set of alternative things) that can coherently be called "things going best," even if this is only a rough characterisation, is to think something much stronger, much more global, than just having a variety of views about particular cases of comparative value. On pain of arbitrariness and/ or debilitating vagueness in its foundations, consequentialism needs this much stronger thought. It cannot have it.

nothing; as Parfit agrees (OWM I: 378), there is not just *one* such scenario. If any such scenario could be described at all convincingly, lots could, and all of them might well dictate quite different principles. The same is true if we alter "things going best" to "things going well in the following specific way. . . ."

Either alteration introduces an important arbitrariness into the foundations of consequentialism. When we talk about "things going best," we are not really, literally, talking about the best conceivable state of affairs (or even the best conceivable state of affairs given unchangeable facts, e.g. the laws of nature and previous history of our world).[15] Unless we go for the third alternative in the last paragraph, we are not even talking, with any clarity or determinacy, about any particular way the world could be. We are talking vaguely and impressionistically about things going *well*. And insofar as we manage determinacy, we are talking about one particular way for things to go well, which apparently we have fastened on pretty much arbitrarily, simply because it strikes us or grabs our attention.

Once it gets in through R2, this arbitrariness infects the argument's third and fourth premises. Strictly speaking, R3 and R4 are now indeterminate in sense because the reference of "these principles" is no longer fixed. Even if we can patch up that problem, there are other objections too. Why would it be true (R3) that everyone could rationally will the set of principles that would make things go best, or (R4) that if there was any set of principles thus willable, there would (even in most cases) be just one such set? Consider this, from Christine Korsgaard:

> [W]hat makes the sum of [everybody's happiness] desirable? Mill wants to mean that each *part* of it is desired, by the person whose happiness it is. But of course a maximum does not include its parts in *that* way: maximising happiness is not like adding one acre of ground to another that adjoins it. Conflicts are possible, and if the calculation turns out so, I may have to sacrifice my happiness in order to maximise the total, and then where is my part?
>
> Someone who challenges the principle of utility when his own happiness is to be sacrificed is not denying that there will be more total happiness if we follow the principle of utility. He is asking why he therefore has a reason to give up his own happiness, which the utilitarian must agree is also a good. The utilitarian can try to block this challenge by insisting that a maximum of happiness is obviously better . . . But the utilitarian cannot base this claim on the idea that a maximum of happiness includes any person's happiness and more, because it doesn't; someone's happiness may be left out. [C. M. Korsgaard, *Self-Constitution* (Oxford UP, 2009), pp. 54, 56–57].

15. Thanks here to Sarah Broadie.

The idea that horrendous sacrifices might be demanded of individuals to secure the overall good – torturing one person to save a million, etc. – is familiar enough from critiques of consequentialism. There is no reason a priori to think that "the set of principles that will make things go best" will not involve such sacrifices. If it does, then there is no reason, either, to think that these sacrifices will be ones that the individuals involved can rationally will.

To deal with this problem, and to make it come out that there *is* a set of principles that everybody can rationally will that everybody should accept, the triple theory needs to set a level of sacrifice to the agent that is consistent with that agent's accepting that sacrifice without irrationality. But then, two consequences follow.

First, it becomes clear that "rational willing" in Parfit's sense is consequentialist willing, the trading-off of costs against benefits. This is neither what Kant had in mind nor, so far as I can see, close enough to Kant to count even as a feasible revisionism about Kant. Parfit is quite open about his revisionary reading of Kant (OWM I: 17): "It may be objected that, if we revise Kant's formulas . . . we are no longer discussing Kant's view. That is true, but no objection. We are developing a Kantian moral theory, in a way that may make progress." But is the theory even Kantian?

In the second consequence, *prima facie*, it seems entirely possible that following the set of principles that we derive by Parfit's methods will look so far suboptimal to a more orthodox consequentialist that he will prefer to follow some policy with higher payoffs, *even though* this policy is not one that everybody can rationally will that everybody should accept. 4 Maybe there is some policy that gives us four times as much utility, but which only 90% of people can rationally will that everybody should accept, or that everybody can rationally will that only 90% of people should accept . . . and so on; complexities reminiscent of Michael Ridge's variable-rate rule utilitarianism.[16] As Parfit himself agrees, anyone consequentialistically inclined might well follow one of these policies rather than the policy that the triple theory proposes.

In the remainder of this critical study, I turn from Parfit's discussion of the triple theory to his method and style, and to a key philosophical issue that they raise. Consider this (OWM I: 256):

> Suppose again that, in
>
> *Rescue*, a hundred miners are trapped underground, with flood-waters rising. These miners will all be saved if four people join some rescue

16. Michael Ridge, "Introducing variable-rate rule utilitarianism," *Philosophical Quarterly* (2006) **56** (223): 242–253.

> mission. I know that four other people have already joined this mission. I could either join this mission as well, or go elsewhere and save the life of some other single person.

On the whole scheme view, I ought to join this mission, since my act will then be one of a set of acts that will together do the most good, by saving a hundred people. This is clearly the wrong conclusion. I ought to save the single person, since one more person's life would then be saved. 5

The argument that Parfit presents here against what he labels "the whole scheme view" is intended to be decisive. But it can only be decisive if we share a whole collection of unspoken but non-trivial assumptions with Parfit, assumptions that serve to close off all practically significant redescriptions of the situation.

For example, Parfit must mean us to assume that I do not know (as in practice I obviously might) that one or more of the four already involved in the miners' rescue is a deeply unreliable person, or someone who is liable to faint, drop dead or sink into a drug-crazed stupor before she/he has done what needs to be done.

Again, Parfit must be assuming that the "other single person" whom "I" can save in *Rescue* is in some way or other already salient to me. It is not that my alternatives are (i) to join the mission to rescue the miners and (ii) to commit myself, instead, to the bizarre project of simply looking for someone – anyone, anywhere – to rescue. [How could (i) be a serious alternative, for anyone remotely sane?] The alternative to (ii) that Parfit has in mind is rather something like (iii) "rescue such-and-such particular individual already known to me, and already in need of rescue." 6

Third, it matters to Parfit's argument that (i) and (iii) *are* my alternatives. (My salient alternatives? My only alternatives?) If I had other options besides (i) and (iii) that we needed to consider, then his argument might well not be conclusive – which is the Latin for "closed off" – in quite the way that he wants it to be.

Fourth, Parfit must also be closing off as illegitimate the possible response to *Rescue* that is summed up by the words "One of these miners is my wife; no matter what else may be going on, I am going to make *absolutely certain* that she is rescued, that she will know that that is what I did, and that I will be there waiting for her when she gets out of the mine." (Where it was applicable, how *could* this response be illegitimate?)

And so on and so forth, without any obvious stopping point. This 7 indefinitely extensive background of shared assumptions is essential to Parfit's argument; but it is not morally insignificant. Its contents are not trivial or irrelevant propositions, like the assumptions that, throughout the time-period in question, "If p, then p," and the law of gravity will continue to apply.

Deliberations about what to do in cases like *Rescue* that are not closed off in this way, which do *not* take for granted non-trivial background assumptions but actively question them, are almost entirely heterodox in contemporary moral philosophy. Yet it is quite possible to imagine non-closed-off – let us call them *open* – deliberations. Indeed, open deliberations are, at least normally, a good deal more natural to us than closed deliberations:[17] as Dickens reminds us, in *Hard Times*, through the confessions of the fairground ragamuffin Sissy Jupe.

> Mr. M'Choakumchild . . . said, This schoolroom is an immense town, and in it there are a million of inhabitants, and only five-and-twenty are starved to death in the streets, in the course of a year. What is your remark on that proportion? And my remark was – for I couldn't think of a better one – that I thought it must be just as hard upon those who were starved, whether the others were a million, or a million million. And that was wrong, too . . . Then Mr. M'Choakumchild said, I find . . . that in a given time a hundred thousand persons went to sea on long voyages, and only five hundred of them were drowned or burnt to death. What is the percentage? And I said, Miss; here Sissy fairly sobbed as confessing with extreme contrition to her greatest error; "I said it was nothing." – "Nothing, Sissy?" – "Nothing, Miss – to the relations and friends of the people who were killed" (Charles Dickens, *Hard Times*, Chapter 9).

What M'Choakumchild finds in the child Sissy Jupe – and labours, indeed, to choke – is a natural propensity for open rather than closed deliberation. In analytic moral philosophy classes all over the world right now, that same propensity is being carefully drilled out of students by their tutors' expositions of trolley problems, cave problems, transplant problems, rescue problems and the rest of the usual applied-ethics diet of hard-case thought experiments. Few philosophers are explicit or self-conscious about this, but Peter Unger is:

> Toward having the puzzle be instructive, I'll make two stipulations for understanding the examples. The first is this: Beyond what's explicitly stated in each case's presentation, or what's clearly implied by it, there aren't ever any bad consequences of your conduct for anyone and, what's more, there's nothing else that's morally objectionable about it. In effect, this means we're to understand a proposed scenario so that it is as boring as possible. Easily applied by all, in short the stipulation is: Be

17. In a sense, such "deliberation" about hypothetical cases is not strictly speaking *deliberation* at all, since it does not lead to real actions. But what it does do is habituate us into ways of approaching cases of real action; in general, I would say, for the worse. (Thanks again to Sarah Broadie.)

boring! [Peter Unger, *Living High and Letting Die* (Oxford UP 1996, pp. 25–26].[18]

Is this a *good* thing that we who teach philosophy are doing to our students? There seems to be a danger that what we are offering them is a training in the failure of their imaginations and of their natural human sympathies. The typical philosophical use of the "thought experiment" in ethics is not just not to take students of ethics in the same direction as they go in when they read fictional narratives, towards wide-ranging, lateral-thinking, unpredictable, creative explorations of the indefinite possibilities of human life and action. It is to take them in exactly the opposite direction: to channel them down an ever-narrowing modal funnel within which all possible readings of a schematically described situation except for one or two are remorselessly eliminated. This is indeed a training to which the injunction "Be boring!" is apposite. And the normal penalty for failing to be boring in the required way is the same for our students as it was for Sissy: it is a fail.

The question of style in moral philosophy will not go away. In other places besides the one just alluded to, the issue of thought experiments that has so often been raised in specific connection with Parfit's work, it will prove itself – as Bernard Williams often predicted[19] – a deep question. "To discover the right style [in moral philosophy] is to discover what you are really trying to do"; "the aim is to sharpen perception, to make one more acutely and honestly aware of what one is saying, thinking and feeling."[20] To assume, as it seems analytic philosophers now standardly assume (whether reflectively or unreflectively), that there is no more an issue about style for philosophy than there is for physics, is to presuppose an answer – a brief and dismissive one – to that question. Insofar as philosophy, especially moral philosophy,[21] is neither physics nor anything like it, I am sure that that presupposed answer must mislead us.

---

18. Unger's other stipulation about his cases is also a be-boring stipulation: it is to hold motives constant (26). "As much as can make sense, the agent's motivation in one contrast case, and its relation to her conduct there, is like that in the other."

19. "The problem of how to find a style in moral philosophy is actually one of the deepest questions about it. For how do you make it realistic without trying to turn into an amateur novelist, or, again, an armchair anthropologist? Or, if you stick to your last and don't try and do those possibly phoney things, how do you stop relapsing into the meticulous boredom which has surrounded the subject?" (Bernard Williams in an interview with Bryan Magee, *The Listener* 4.2.71, pp. 136–140, at p. 140)

20. This "quotation" is a *mélange* of two passages from Bernard Williams' *Morality: an introduction to ethics* (Cambridge UP 1972) – respectively from p. 11 in the original edition (in the original Preface), and from p. xv of the new Preface that Williams wrote for the Canto reprint of *Morality* (Cambridge UP 1993).

21. It is of course true that different relations to physics, and to science more generally, apply in different parts of philosophy. I have no intention of denying that, but my focus here is ethical.

It must be obvious by now that Parfit can have few readers who are less ideologically sympathetic than I am to his project and his method in OWM. (Or at least to the parts of his project I have discussed here. On some other matters, such as opposition to moral irrealism and to desire-based accounts of practical reason, which Parfit tells me are at least as much his main purpose in OWM as the development of the triple theory, he and I are very much on the same side. Perhaps I have displayed another pattern of regrettably typical academic behaviour by focusing on my *dis*agreements with Parfit.) Given our ideological differences, it is a mark of the excellence of his book how much even I can get from reading it. In his initial introduction of Kant's philosophy, on p. xliv of volume I, Parfit delightfully suggests that the Oxford mark for Kant's work might be an Alpha/ Gamma. If as I suspect something similar is true of Parfit's other master's greatest work, Sidgwick's *Methods of Ethics*, and indeed of this book too, then Parfit finds himself in good and enduringly valuable philosophical company; perhaps the best kind of philosophical company there is.[22]

22. For valuable discussions of Parfit's book over several years, and/or of this critical notice, I am grateful to Sarah Broadie, John Cottingham, Roger Crisp, Garrett Cullity, Hallvard Fossheim, John Haldane, Edward Harcourt, Brad Hooker, Gerald Lang, Adrian Moore, Tim Mulgan, Alan Thomas, and above all to Derek Parfit himself, who sent me more than one enormous preprint of the book, responded patiently and in detail to a long series of objections to its argument from me that he can hardly have had much sympathy with, and even found time to send me – almost by return email – 11 pages of comments on an earlier draft of this critical study.

Toppan Best-set Premedia Limited	
Journal Code: PHIN	Proofreader: Elsie
Article No: 1470	Delivery date: 29 January 2012
Page Extent: 15	Copyeditor: Angela

AUTHOR QUERY FORM

Dear Author,

During the preparation of your manuscript for publication, the questions listed below have arisen. Please attend to these matters and return this form with your proof.

Many thanks for your assistance.

Query References	Query	Remark
q1	AUTHOR: Please provide the full first name for the authors running head.	
q2	WILEY-BLACKWELL: Please confirm if article type is correct.	
q3	AUTHOR: Please provide the full first name for the author.	
q4	AUTHOR: In the second consequence . . . accept. This sentence has been reworded for clarity. Please confirm if this is correct.	
q5	AUTHOR: On the whole scheme view . . . saved. Please confirm that this paragraph has been set correctly.	
q6	AUTHOR: The alternative to (ii) that . . . rescue. In this sentence, please confirm if "(ii)" is correct.	
q7	AUTHOR: And so on and . . . point. Please confirm if this sentence is correct.	

q8	AUTHOR: Please check this website address and confirm that it is correct. (Please note that it is the responsibility of the author(s) to ensure that all URLs given in this article are correct and useable.)	