London Review of Books

Five Girls on a Rock

Allan Gibbard

On What Matters by Derek Parfit
Oxford, 540 pp. and 825 pp, £30.00, May 2011, ISBN 978 0 19 926592 3

Morality can't just be a system of arbitrary taboos. We want its protections, and others want those same protections against us. A morality worth heeding must have a rationale. A chief task of moral philosophers is to discern such a rationale and to shape it by criticism and argument. Derek Parfit's *On What Matters* looks to two great moral philosophers, Immanuel Kant in the late 18th century and Henry Sidgwick, whose treatise *The Methods of Ethics* first appeared in 1874. Kant, Parfit writes, 'is the greatest moral philosopher since the ancient Greeks', but Sidgwick's *Methods* 'is, I believe, the best book on ethics ever written'. Kant and Sidgwick are normally taken to stand for the two great opposing moral visions: Sidgwick for utilitarianism, which concerns itself with how to maximise happiness, and Kant for a moral law grounded in reason. Parfit finds, however, that Kant and Sidgwick are 'climbing the same mountain' by different routes. We are still far from the summit by either route, but as Parfit said a quarter-century ago, 'compared with the other sciences, non-religious ethics is the youngest and the least advanced.' As with any science, a mature ethics might take generations to formulate.

'Actions are right,' the utilitarian John Stuart Mill wrote, 'in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness.' At times, however, duty and the general happiness seem to be in conflict. If a recently dead husband had been leading a secret life, then telling the widow a sweet lie may forestall even greater misery, but if she really wants to know the truth, no matter how bitter, the lie wrongs her. Utilitarianism can be at odds with our moral intuitions. Sidgwick's 'great, drab book', as

Parfit calls it, argued nevertheless that 'the morality of common sense is unconsciously utilitarian.' Sidgwick agreed with his philosophical opponents that moral knowledge must rest ultimately on intuition, but insisted that when intuition is rendered coherent, it supports utilitarianism.

When Parfit and I were young in the 1960s, utilitarianism or something close to it was widely accepted among leading moral philosophers. Parfit developed his astoundingly original early ideas in *Reasons and Persons* (1984), where he examined the logic of alternative forms of utilitarianism and advocated concern for others by a set of Buddhist-like arguments that there need be no clear boundary between others and oneself. *Reasons and Persons* may be the most important work in moral philosophy of its decade.

Over the years, however, moral philosophers have largely come to reject utilitarianism. The conflicts with our moral intuitions seem just too damning. Suppose, in a hospital, five patients will die unless organs can be found for transplants, and a healthy young man comes in for a check-up. Cutting him up for his organs would maximise the total happiness, but to do this would clearly be wrong: it would violate his right to life. So much the worse, then, for utilitarianism. What sort of alternative rationale would spare him? A century before Sidgwick, Kant offered an answer. Act only, Kant's Categorical Imperative commands, by that maxim you can will at the same time to be a universal law. The spur to morality is reverence for this law. We must always respect the rational nature that makes us human. Kant's formulas, many philosophers think, explain why our intuition abhors the thought of killing the young man for his organs, however greatly they may be needed. To do so without his consent fails to respect his rational nature. Just how Kant's alternative to utilitarianism operates has been a matter of contention, but the conviction has grown that he did indeed find an alternative basis for morality, one to which moral intuition responds.

On first inspection, though, it isn't clear that Kant's formulas really do exclude utilitarianism. He himself, it's true, despised utilitarianism as he knew it, but do his arguments really tell against it? Parfit's teacher, later colleague, R.M. Hare, who dominated Oxford moral philosophy in the later part of the mid-20th century, proclaimed himself a Kantian, but argued that what is coherent and systematic in Kant turns out to be utilitarian. Most philosophers who follow Kant have regarded Hare as an oddball and his Kantianism as bogus. But maybe Hare has a point. If we cut up the young man for his organs, how have we violated any of Kant's dicta? Perhaps the fault consists in failing to respect his humanity. How, though, would letting five others die when we could have saved them count as respecting *their* humanity? By killing the young man for his organs, some Kantians say, we treat him

merely as a means, as a thing and not as a rational being. But Kant never said not to treat people as means: we do so whenever we call on their help. Kant's demand is that we always treat rational nature as an end. In this case, we do just that: we weigh the good of the young man into our decision along with everyone else's. Some Kantians take a different approach. The problem, they say, is that he couldn't consent to be sacrificed. Yet if we spare him and let the others die, do they consent to this neglect? Whatever we do, someone isn't going to survive, and we must treat all six individuals as ends. We respect people, another Kantian might say, by living with them on the basis of a social contract we would each of us have ratified. Why, though, wouldn't we ratify a system that, in such tragic cases as this, maximises everyone's chances? It might have been the young man who needed a transplant that could be had only by cutting up someone else.

These retorts may have answers, but to be fair to Hare, some of the answers are equally available to utilitarians, most of whom would also reject killing the healthy young man. Survival unscathed, a utilitarian can say, is better than survival via a transplant, and a system that allowed murder for the harvesting of needed parts would spread fear and mistrust. Some utilitarians concede that cases like this are a problem for utilitarianism in one of its forms: they tell against act-utilitarianism, the doctrine that the right course of action is one which maximises the net happiness of all affected. These philosophers instead adopt rule-utilitarianism: what makes an act right, they say, is that it accords with happiness-promoting rules, the rules whose general acceptance would make for the greatest net happiness. General acceptance of a rule allowing us to kill people for their body parts would spread terror and foster abuse, and that is why killing the young man would be wrong.

Kantians who oppose utilitarianism offer other explanations besides the ones I have mentioned. Many of them, however, just presuppose a moral asymmetry between killing a person and letting him die. Intuitively, we for the most part abhor killing more than we oppose letting someone die, but if Kantian formulas are to provide a deep explanation of why our intuitions tend to get morality right, they must explain this asymmetry and not just help themselves to it. We each have a right not to be killed, but not an equally imprescriptible right to be saved, so it won't count as disrespectful to fail to save a person when the only way to save her is to kill someone else. How, we need to ask, does any of this emerge from the deeper basis for morality that we find in Kant?

My treatment of these worries has been quick and rough. Parfit's inquiry into these and many other issues covers hundreds of pages. Sometimes he elaborates from Kant's starting points; sometimes he finds a Kantian formulation empty or wild in its implications, and revises it or

moves on to alternatives. Kant appeals, for example, to the respect we should feel for rational nature, but these appeals 'add little to Kant's view'. Kant 'writes that any liar "violates the dignity of humanity in his own person"', becoming a 'mere deceptive appearance of a human being' who has 'even less worth than if he were a mere thing'. But 'these are not the claims,' Parfit judges, 'that make Kant the greatest moral philosopher since the ancient Greeks.' What does stem from genuine moral insight, leading us to a systematic ethics, is the demand Parfit calls Kantian contractualism. As he formulates it, 'everyone ought to follow the principles whose universal acceptance everyone could rationally will.' This, Parfit thinks, may be the true grounding principle of morality.

To illustrate his main argument, Parfit gives an example stripped of all the usual complexities. Five girls are stranded on a rock in the sea, while on a different rock a sixth girl is stranded alone. I have a lifeboat and can save the five or save the one, but not both. All are strangers to me, and each, if she survives, will lead a life equally worth living. Utilitarians tell me to save the five, but what would Kant do? He tells us to treat each girl as she could rationally will. Now each could rationally will to be saved herself, even if it costs the lives of all five others. But the lone girl, Parfit thinks, might also rationally will the five others to be saved, even at her own expense. Reason doesn't require this, but it does allow it. Saving the five, then, is something each of the girls could rationally will. And since there is nothing else that each of them could rationally will – it would be crazy for one of the five girls to will herself and four others to die to save one – saving the five is the only thing that all could rationally will, and so morally it is what I must do. In this case, Parfit concludes, utilitarianism and Kantian contractualism coincide.

Parfit takes this pattern of argument and generalises it. It's not precisely utilitarianism that he derives from Kant. Utilitarians say two things: first, that what's good in itself is happiness, and second, that we ought to promote the good. Parfit's aim is to establish the second but not the first. The doctrine that we ought to maximise the total good is known as consequentialism. Utilitarianism is one form of consequentialism, but a consequentialist may also hold that things other than happiness are good in themselves: knowledge, for example, or accomplishment, genuine friendship or social fairness.

In these volumes Parfit doesn't try for a theory of what's good, but even so his consequentialism has teeth. By his definition, a good is something that, from an impartial standpoint, we have reason to want. Abstractly, then, a consequentialist is someone who says we should promote whatever it is we have impartial reason to favour. Parfit is a rule-consequentialist, not an act-consequentialist. In the transplant case, the test of rightness is

whether the world would be better if there were general acceptance of rules that would allow the killing of the young man. It presumably wouldn't. (No one in these debates really advocates killing people for their organs. The debates are over how to explain what's wrong with doing so.)

Does Parfit, then, succeed in his grand argument that Kant and Sidgwick are both heading towards the same summit? The respondents to his Tanner Lectures at Berkeley in 2002 – *On What Matters* includes revised and expanded versions of these lectures and the respondents' commentaries – maintain that he doesn't. I think he succeeds beautifully, but I may be a soft touch: I argued in my own Tanner Lectures that contractualist and utilitarian approaches both lead to some form of consequentialism if they lead anywhere. What do Parfit's critics have to say against this? He leaves out, they charge, what Kant thought key to morality. One of them, Barbara Herman, speaks of a 'mismatch of methods'. With this Parfit agrees: he insists that his theory is genuinely Kantian, in that it refines insights and doctrines that are Kant's – but they are not the ones Kant himself believed fundamental. To act morally, Kant was convinced, is to be consistent in deciding what we should do. What makes wrong acts wrong is their incoherence – as, for example, when they make an exception for ourselves. Consistency in action, furthermore, demands respecting above all a person's capacity to decide rationally what ends to pursue. These insights are central to Kant's theorising, but Parfit believes they are mistaken. He tries out other potential moral principles too, one after another, but finds them all either incoherent or vastly implausible.

The commentators included in *On What Matters*, so far as I can see, don't exactly reject this assessment. Herman clings to a hope that morality can be grounded as Kant contended, but leaves open the question as to whether it can. Susan Wolf thinks that morality has no one supreme principle, but that Kant identifies moral considerations of prime importance – autonomy above all – that compete, at times tragically, with other moral considerations such as welfare. Allen Wood objects to the use of lifeboat cases as tests, since they leave out the social and institutional factors that ordinarily matter to us and on which our moral intuitions are trained. Still, he finds, there are cases in which Kantian ways of thinking can't tell us what to do, and then our moral aspirations simply can't be met. None of Parfit's respondents claims there is a truly Kantian alternative to what he is attempting.

The remaining commentator, T.M. Scanlon, has his own form of contractualism: the morality of what we owe to each other consists in living with others on principles 'which no one could reasonably reject as a basis for informed, unforced general agreement'. This shares a good deal with Parfit's reading of Kant, but Parfit's contractualism isn't Scanlon's. For what we owe

to each other, Scanlon maintains, numbers don't count. In the lifeboat case, you rescue more girls if you rescue the five, but you don't on that account owe them rescue. Parfit finds this implausible, and so do I. What Parfit shows, if his arguments are successful, is that there isn't a distinct, coherent Kantian alternative to consequentialism that could underlie our moral intuitions. Kant's slogans often sound as if they support something different, but mostly they don't, or the alternatives they support are grossly implausible in their implications.

Why, though, heed moral intuition at all? This is a question of metaethics, a term for the parts of ethical theory that examine what moral claims consist in and how we can have moral knowledge. Parfit devotes nearly half of these two thick volumes to metaethics. Fundamental in ethics, he holds, is the concept of a *reason* to do something or want something. Is there reason to want those who hurt you to suffer, even apart from any further good that their suffering might produce? This isn't meant as a question about why people do the things they do – their reasons *for* doing things – but about reasons *to* do things: 'normative' reasons, in philosophers' jargon. Not all reasons are moral. Hitting your own thumb with a hammer would cause you great pain, and clearly that's reason not to do it – a non-moral reason. Cases like these are obvious: our intuitions about them are clear and unconflicted. I also have reason not to hit your thumb with a hammer, and you have reason not to hit mine, and most of us would find this almost equally obvious. That my hitting your thumb with a hammer would cause you pain is reason not to do it – a moral reason. The problem for moral theory is how to reach beyond such findings to the cases we find puzzling.

Parfit's theory of what reasons are is a form of non-naturalism. That there is reason not to torment oneself or others isn't a natural fact: it is not, that is to say, the sort of fact we can confirm by observation, as we can with the facts of physics or psychology. If something is a reason, that's a fact, a fact that is not purely natural. As for how we can know such facts, Parfit holds that most fundamentally we know about them by intuition. With his non-naturalism and intuitionism, Parfit agrees with his hero Sidgwick (my hero too). Two decades ago, most ethical theorists saw the non-naturalism that Parfit cleaves to as a relic of a more credulous age, but more recently a host of ethical theorists have embraced it. A philosophical reader is likely to find Parfit's arguments definitive against every position but one: whichever argument convinces that reader. I find what he says convincing in almost every respect, yet find the position he ends up with impossible to believe, and some of his attacks unconvincing (including some sharp attacks on writings of mine).

According to Parfit, we come to know non-natural facts when we reach reflective equilibrium: a state in which a conviction can survive any possible criticism. Why, though, would believing

a thing in a state of reflective equilibrium be a sign that it's so? Even without an answer to that question, we can proceed with our thinking more or less in good conscience. The sciences, after all, often pin down their results long before anyone finds a full and coherent story to explain them. Something like reflective equilibrium must be what we seek with any fundamental and puzzling philosophical issue. Parfit, though, isn't just getting on with his moral thinking and postponing questions of what it consists in. Rather, over many, many pages, he explains the special nature of his subject, and it is therefore fair to ask whether the story he tells is coherent and defensible. Are non-natural facts credible? If we put our moral intuitions into reflective equilibrium, have we thereby discovered non-natural facts?

Parfit cites mathematics. Pure mathematics, too, is non-natural and known by pure thinking, not by observation. Ultimately, it is by intuition that we know what must hold mathematically. When our intuitions conflict with one another, we think about how to make them consistent and move towards reflective equilibrium. If non-natural facts seem fishy, the alternatives are even worse – as Parfit shows by assembling familiar arguments and devising new ones. Take the claim: 'If a donation would prevent starvation, that is a reason to donate.' This means, according to some philosophers, that if one vividly and repeatedly considered the natural facts involved, and believed that donating would prevent starvation, one would be motivated to donate. That one would be motivated to donate is a natural fact – a psychological fact. But this fact is different from the fact that there's reason to donate. Even if one wouldn't be so motivated, it is intelligible to claim that there is reason to donate: we understand what such a claim would mean. (Parfit reports that Bernard Williams seemed genuinely not to understand such claims, but he thinks that may be because Williams lacked a concept that the rest of us have.) Parfit argues in a similar vein against a wide range of positions in which all genuine facts are held to be natural. Non-naturalism, Parfit concludes, is what's left standing; it is also, he claims, the theory that answers best to human concerns. How do we know a non-natural fact? When we 'see', as it were, that a mathematical claim must be true, we are 'responding in non-causal ways to the validity of some kinds of reasoning'. Likewise with reasons: we respond non-causally to facts of what basic reasons there really are.

One test of any such claim is whether it fits in with the best accounts we have of thinking and judgment. Our brains are an outcome of Darwinian natural selection and other natural processes, and the truth about reasons and how we know them must somehow fit in with this naturalistic story. How, then, could natural processes have equipped us to know non-natural facts? Our cognitive abilities, Parfit agrees, 'were partly produced by evolutionary forces', but 'these abilities later ceased to be governed by these forces, and had their own effects.' The

ability to reason 'is sometimes claimed to be mysterious. But when it seems to us clear that some belief must be true, there is nothing in our cognitive experience that is more transparent and intelligible, or less mysterious.'

It's true enough that we are not caused to be what we are by genes alone, yet in some way, patterns of neural goings-on constitute our making the judgments we do, and the ways these patterns come into being make them reliable as knowledge or not. There is much that is mysterious in all this, no matter how 'transparent and intelligible' we experience it as being. How could non-causal responses to non-natural facts be part of any such story? What does 'responding' non-causally even consist in? Our brains don't work by magic, and a satisfactory story must indicate how mathematical and normative judgments might operate. I have some idea why intelligent, evolved organisms would get arithmetic and geometry right: carpenters who bungle geometry mess up their work, and accordingly those who got such things right tended to reproduce more. Ethical judgment may affect rates of reproduction – it is plausible, for example, that reproductive success would be enhanced by the eliciting of co-operation on mutually advantageous terms – but how could fitting the normative facts be part of the story?

Parfit has more to say on these matters, and I don't take myself to have refuted him with these few words. His magnificent compendium of arguments, though, leads him to a position that we don't know how to reconcile with the view that human phenomena are natural – and with each passing decade we get further indications that naturalistically is the way we must learn to understand ourselves. Still, pending a credible and comprehensive view of right and wrong and our powers to know the difference, we don't really have an alternative to Parfit's way of approaching substantive moral questions. Something like reflective equilibrium is what we're stuck with for now, trusting our judgments discriminately and critically and seeing where it all leads. We don't know any way to do systematic moral thinking that doesn't amount to relying on those of our judgments that stay plausible under severe testing.

It might seem worrying that different people could engage in such an inquiry yet come to opposed conclusions. One of Parfit's principal aims is to remove any such anxiety. 'If everyone knew all of the relevant non-normative facts, used the same normative concepts, understood and carefully reflected on the relevant arguments, and was not affected by any distorting influence, we and others would have similar normative beliefs.' Even Nietzsche would join in, Parfit argues in a fascinating chapter.

I myself am puzzled whether to accept the intuitions that ground Parfit's arguments for moral convergence. I find these intuitions plausible but far from compelling. He himself claims no

knock-down argument. Rather, to my mind, the lesson of his inquiry is how very difficult it is to get plausible moral thinking to be systematic unless it eventuates in some sort of consequentialism. Attempts to get systematic moral thinking to be non-consequentialist lead to extravagances and special pleading; Scanlon's insistence that numbers don't count strikes me as one instance of this.

If there were no non-natural properties, Parfit tells us at one point in a tone of despair, then nothing would matter. For something to matter, after all, is for there to be reason to care about it, and facts about reasons are non-natural. So mattering, Parfit insists, must be non-natural. But if no properties are non-natural, what follows is an either/ or: either nothing matters or Parfit is wrong that mattering is non-natural. Now it seems beyond doubt that things matter, whereas we needn't wholly trust in Parfit's metatheory of mattering. If there are no non-natural properties, we should conclude, things still matter and Parfit is wrong about what mattering is.

It matters what matters. Whether mattering is non-natural or not matters too, but not as much. Either way, it matters considerably whether Parfit's basic methods of discerning how things matter are reliable. I am uneasy as to whether they are, but I see no alternative. Still, if these methods are reliable, non-naturalism leaves this reliability unexplained. I keep hoping that we will come to understand questions of what matters in a way that yields more insight than non-naturalism offers.

Vol. 34 No. 11 · 7 June 2012 » Allan Gibbard » Five Girls on a Rock
pages 22-24 | 4184 words

 ^ Top