Review of Parfit, *On What Matters*

Allan Gibbard

Morality can’t just be a system of arbitrary taboos. We want its protections, and others need those protections from us. A morality worth heeding must have a rationale, and considerations like these must somehow drive it. A chief job of moral philosophers is to discern such a rationale, and to shape it by criticism and argument. Utilitarians Bentham and John Stuart Mill grounded morality in the general happiness—but at times, duty and total happiness seem to conflict. If a dead husband had led a secret life, then a sweet lie told to the widow may forestall even greater misery, but if she would really want to know the truth however bitter, the lie wrongs her. Utilitarianism is sometimes at odds, it seems, with our moral intuitions.

Henry Sidgwick’s *The Methods of Ethics* in 1874—“his great, drab book”, as Derek Parfit calls it—argued over hundreds of pages that whereas moral knowledge must rest ultimately on intuition, “the morality of common sense is unconsciously utilitarian.” Utilitarianism, in Sidgwick’s version, is the doctrine that morally right acts are those that produce the greatest total happiness for all. When Parfit and I were young in the 1960’s, utilitarianism or something close was widely accepted among leading moral philosophers. Parfit developed the astoundingly original ideas of his youth in his vast 1984 book *Reasons and Persons*. His line was more or less utilitarian, scrutinizing with avid precision the logic of alternative forms of utilitarianism and supporting concern for others by a set of Bhuddist-like arguments that there need be no clear boundary between one person’s experience and another’s. That book may stand as the most important book in moral philosophy of its decade.

As he and I grew older, though, moral philosophers came widely to reject utilitarianism. Conflicts between utilitarianism and our moral intuitions just seemed too damning. A healthy young violinist visits a hospital where five patients will die unless organs can be found for transplants. Cutting him up for his organs maximizes total happiness, but doing so would clearly be wrong. It would violate his right to life—and so much the worse, then, for utilitarianism. What sort of alternative rationale for morality would spare him? From a century before Sidgwick comes Immanuel Kant with answers. Morality, is grounded not in happiness but in reason. The spur to morality is reverence for the law of action which commands us always so to act that we could will our maxim as universal law. Always we must respect humanity, the rational nature that makes us human. To violate the young violinist’s rights by harvesting his organs, however greatly they may be needed, fails to respect his rational nature. In Nietzsche’s aphorism, “Kant wanted to prove, in an everyman-head-buffeting way, that everyman was right.”
Just how Kant’s alternative to utilitarianism works has been contentious, but the conviction has grown that he finds an alternative basis for the morality we glimpse with our moral intuitions.

On first inspection, though, it isn’t clear that Kant’s formulas really do exclude utilitarianism. Kant himself despised utilitarianism as he knew it, but do his arguments really condemn it? Parfit’s teacher and colleague R. M. Hare, who dominated Oxford moral philosophy in later mid-century, proclaimed himself a Kantian, but argued that what’s coherent and systematic in Kant turns out just to be utilitarian. Still, most philosophers who adhere to Kant have regarded Hare as an oddball and his Kantianism as bogus. If, though, we cut up the violinist for his organs, how we have violated Kant’s systematic dicta? Does violating his right to life fail to respect his humanity? Why would letting five others die when we could have saved them count as respecting their own humanity? By killing him for his organs, it might be argued, we treat him merely as a means, like a thing and not as a rational being. Still, we treat people as means whenever we call on their help—and Kant never forbade this; to do so would be preposterous. His requirement is that we must treat rational nature also as an end. In this tragic situation, we do just that: we weigh the young violinist’s good into the decision along with everyone else’s. That is not how we treat things. The problem, we could try saying, is that he couldn’t consent to the way we treat him. If we spare him, though, and let the others die, could they consent to be neglected? Whatever we do, someone won’t survive, and we must treat as ends all six. We respect people, we could try saying, by living with them on the basis of a social contract we each would have ratified. Why, though, wouldn’t we agree on a system that, for tragic cases, maximizes our chances? It might have been the violinist who needed a transplant that could be had only by cutting up someone else. These objections may have answers, but some of these answers are available just as much to utilitarians. Survival unscathed is far better than survival with a transplant, and a system that allowed murder for needed parts would have all sorts of terrible side-effects. Other responses already presuppose the asymmetry between killing a person and letting him die, something that Kant is called on to explain. True enough, it won’t count as disrespectful to fail to save a person when the only way to save her is to kill someone else. That, however, is because we each have the right not to be killed, but not an equally imprescriptable right to be saved. How does any of this emerge from a deeper basis that we find in Kant?

My rendering of these worries has been quick and breezy. Derek Parfit, in his remarkable treatise On What Matters, pursues such issues over hundreds of pages. Parfit developed this two volume work from Tanner Lectures he gave at the University of California, Berkeley. The lecture events at Berkeley include commentaries from others and responses from the lecturer, and the lectures, commentaries, and responses are subsequently revised for publication, often in expanded form. The results are rarely bulky (my own such book totals 200 pages), but for eight
long years Parfit made it his major project to rethink and expand his lectures and responses into this work of over 1400 pages.

Kant, writes Parfit, “is the greatest moral philosopher since the ancient Greeks,” but “Sidgwick’s Methods is, I believe, the best book in ethics ever written.” Peter Singer has hailed Parfit’s new work in turn as “the most significant contribution to moral philosophy” since Sidgwick. Simon Blackburn, whose views on many matters are close to mine, responds with an impassioned critique that the Financial Times refused to print.

Kant’s insights yield new bases for morality, Parfit finds, but not bases genuinely at odds with utilitarianism. Kant and Sidgwick are “climbing the same mountain” by different routes toward the summit. We are still far from the summit by either of these routes, but as Parfit said a quarter-century ago, “Compared with the other sciences, non-religious ethics is the youngest and least advanced.” As with any science, a mature ethics might take generations.

How, then, can we work toward a science of ethics? To start with, Parfit looks to Kant. Sometimes he elaborates on Kant’s starting points and develops them; 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 respect for rational nature, but these appeals “add little to Kant’s views.” Kant, Parfit tells us, “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’. These are not the claims”, Parfit judges, “that make Kant the greatest moral philosopher since the ancient Greeks”. Another of Kant’s demands, though, does stem from genuine moral insight and can lead us to a systematic ethics. This Parfit calls Kantian contractualism: as Parfit formulates it, “Everyone ought to follow the principles whose universal acceptance everyone could rationally will.” This may indeed be the true grounding principle of morality.

To elucidate his argument, Parfit strips an example of all usual complexities. On a rock in the sea five girls are stranded, and on another rock is stranded a girl alone. I have a lifeboat and can save the five or save the one, but alas not both. All are strangers, and each, if she survives, will lead a life equally worth living. Utilitarians tell me to save the five; what would Kant do? In rough terms, 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 I can’t save all six. The lone girl, though, Parfit thinks, also might be rational if she self-sacrificingly willed saving the five others, even at the expense of saving her. Not that reason requires this, but it permits her so to will. Saving the five, then, is something each of the girls could rationally will—and since there is nothing else that each could rationally will, this is what I morally must do. Utilitarianism and Kantian contractualism for this case coincide.
Parfit takes this pattern of argument and generalizes it. It’s not precisely utilitarianism that he derives from Kant. Utilitarians say two things: what’s good in itself is happiness, and we ought to promote the good. Parfit’s aim is to establish this second but not the first. This less committal doctrine, that we ought to maximize total good, is called consequentialism. Utilitarianism is one form of consequentialism, but a consequentialism may hold that things other than happiness are good in themselves. Knowledge, accomplishment, genuine friendship, and fairness in social arrangements are candidates.

Parfit doesn’t in these volumes try for a theory of what’s good, but even so, his consequentialism has teeth. By definition, to be good is to be something that, from an impartial standpoint, we have reason to want. Abstractly, then, a consequentialist is one who says to promote whatever it is we have impartial reason to favor.

The promotion, Parfit says, is to be indirect. Parfit is what is called a rule-consequentialist. Is it right to transplant the violinist’s organs without his consent? That’s a question of whether accepting rules that allow such an act would make the world go best. (It presumably wouldn’t. No one in these debates really advocates cutting people up involuntarily for their organs. If someone does, we can recoil or we can debate, but Parfit takes it that of course killing the violinist for his organs would be wrong. He asks what would make it wrong. Accepting such rules would breed fear and abuse.)

Does Parfit, then, succeed in his grand argument that Kant and Sidgwick lead toward one summit? The respondents maintain that he doesn’t. I myself think he succeeds beautifully, but I may be a soft touch for arguments like Parfit’s. My own Berkeley Tanner Lectures argued too that contractualist and utilitarian approaches reach similar endpoints. Parfit and I agree that any one of these starting points, when rendered coherent and developed in ways that aren’t grossly implausible, leads to some form of consequentialism if it leads anywhere. What do anti-consequentialist Kantians have to say against this? The commentaries and replies extend over hundreds of pages. Two of the commentators are pretty much card-carrying Kantians, and the other two agree that Parfit doesn’t draw from Kant all that he should.

Parfit, the critics charge, leaves out what Kant thought key to morality. Barbara Herman speaks of “a mismatch of methods”. With this charge Parfit agrees: Kant’s most fundamental attempts to explain morality aren’t what we can draw from him. Morality, Kant was convinced, is consistency in deciding what to do. Policies of wrongdoing contradict themselves by making an exception of the dear self. Always, what makes wrong acts wrong is their incoherence. Consistency in action, Kant thought further, demands respecting above all a person’s capacity to decide rationally what ends to pursue. These seeming insights are central to Kant’s theorizing, Parfit agrees, but mistaken. Parfit tries out version after version of candidate principles along these lines, and finds them incoherent or vastly implausible. Some of Kant’s insights were of
first importance, he maintains, but they aren’t the ones Kant himself thought most fundamental. Parfit insists that his theory is genuinely Kantian, in that it refines insights and doctrines that are genuinely Kant’s—but they are not the ones Kant himself believed to be deepest.

The commentators, as far as I can see, don’t exactly reject this assessment. Barbara Herman clings to hopes that morality can be grounded as Kant contended, but she leaves it open whether these hopes can be realized. 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 like welfare. With no supreme principle to adjudicate among these considerations. Allen Wood objects to lifeboat cases as tests, cases that excise the social and institutional factors that ordinarily matter and on which our moral intuitions are trained. Still, he finds, there are cases for which Kantian ways of thinking can’t tell us what to do, and then, our moral aspirations just can’t be met. None of the commentators, in short, claim a truly Kantian alternative to what Parfit attempts.

T. M. Scanlon—to whom Parfit dedicates the volumes along with Thomas Nagel—adopts the title “Why I am not a Kantian”. Like Parfit, he rejects Kant’s aspirations to derive morality from the sheer logic of action, and founds moral theory on intuitions about reasons. Scanlon, Parfit argues, offers a third starting point toward the summit of moral theory, a form of contractualism but different from Kant’s. Morality consists in living with others on principles “that no one could reasonably reject as a basis for informed, unforced general agreement”\textsuperscript{153}. This grounds not all morality, according to Scanlon, but a central part, the part that concerns what we owe to each other. Scanlon does not, in his commentary, dispute Parfit’s derivation of Rule Consequentialism from Kantian Contractualism, but Parfit’s “Kantian” 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. I myself find this worrying if taken seriously. You do owe rescue to the five girls, at least if you can save them at moderate cost and low risk to yourself, and you don’t owe rescue to the lone stranded girl, since you could save her only by failing to save the five.

Scanlon’s theoretical motive seems chiefly to be not to kill the violinist. Utilitarianism, he has long maintained, ignores the “separateness of persons”, as Rawls called it. It holds that one person’s good may compensate morally for another’s ill. Parfit’s derivation itself has no such premise, and Scanlon has long recognized too that for all we know at the start, we might forgo any such premise still come to a utilitarian standard of right and wrong. Still, Scanlon wants to save the violinist. He thinks that to satisfy our intuitions, we must hold that numbers don’t always count. Now everyone in these debates wants to rescue the poor violinist from moral theory. The issue is on what grounds. Scanlon, it seems to me, must go into moral contortions to achieve the right conclusion for such cases. Surely, the resources I owe in the end to one
person can be affected by what I owe to others. A sufficient mass of competing obligations can keep me from owing rescue to a girl whom I could save only by letting five others drown. I owe rescue to the five whom I can save together. No one, Parfit and I agree, could reasonably reject a principle for saying this. Why, then, not sacrifice the violinist? Because accepting a rule that allowed it would have terrible consequences.

Kant sets out to find “the supreme principle of morality”, and Parfit joins him in this search. Kantians who reject consequentialism, to judge for these exchanges, aren’t going to further this project. What Parfit shows, if his arguments are successful, is that there isn’t a distinct and plausible Kantian alternative to consequentialism. Kant’s slogans often sound as if they support something different, but mostly they don’t, or the alternatives they support are ludicrous. So argues Pafit at meticulous length.

Why, though, heed moral intuition? This is a question of metaethics, as the parts of ethical theory are called that don’t seek the standard of right and wrong and its bases, but examine what moral claims consist in and how we can have moral knowledge. Chiefly it is metaethics that Parfit adds to the Tanner Lectures’ format of lectures, commentaries, and replies, and he spends well nigh these two thick volumes on it. Conceptually at base in ethics, he holds, is the concept of a reason to do a thing or want a thing. Is there, for instance, 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 in fact 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 put you in agony, and clearly that’s reason in itself not to do it—a non-moral reason. Cases like these are obvious; our intuitions on them are clear and unconflicted. The rest of us too have reason not to hit your thumb with a hammer. Most of us would find this almost equally obvious. For us, that hitting your thumb with a hammer would put you in agony is reason in itself not to do it—a moral reason. These are the easiest of cases for a theory of rationality and morality. The problem for moral theory is how to reach beyond these obvious findings to the cases we find puzzling.

Parfit’s theory of what reasons are is a form of nonnaturalism: that something is a reason is a fact—a nonnatural fact. As for how we know such facts, he propounds intuitionism: most fundamentally, we know about reasons by intuition. I agree with much of what Parfit says in support of these metatheoretical claims, and disagree with some of it. (Among Parfit’s sharp attacks are a number on writings of mine.) A philosophical reader will likely find Parfit’s arguments definitive against every position but one: whatever one convinces that reader. I myself find what he says convincing in almost every respect, and find the position he ends up with impossible to believe—and some of his attacks unconvincing. Two decades ago, most ethical theorists saw the kind of nonnaturalism Parfit embraces as a relic of a more credulous age.
More recently, a host of ethical theorists have embraced this kind of position. It may be the coming thing in metaethics, and a substantial party of metaethical theorists are already on board. Peter Singer is editing a collection of critical pieces on Parfit’s metaethics, and most of what I have to say I’ll leave for then. Here, though, I’ll say a little.

With his nonnaturalism and intuitionism, Parfit agrees with his hero Sidgwick (my hero too). We come to know nonnatural facts if we reach reflective equilibrium, a state where a conviction would survive any possible criticism. Why, though, would believing a thing in reflective equilibrium be a sign that it’s so? Even without an answer, we can proceed with our thinking in more or less good conscience. Sciences, after all, often pin down their results long before anyone finds any full and coherent philosophical story to explain them. Something like reflective equilibrium, we can say at this point, is what we seek with any fundamental and puzzling philosophical issue whatsoever. Parfit, though, isn’t just getting on with moral thinking and postponing questions of what it consists in. Rather, over hundreds of pages, he explains the special nature of his subject. It is fair, then, to ask whether the story he tells can be made 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 intuitions conflict, we consider alternative ways we could render them consistent and move toward reflective equilibrium. If nonnatural facts seem fishy, the alternatives are even worse—so Parfit devotes himself to showing. Problems with alternatives are familiar to any philosopher, and nonnaturalism, Parfit concludes, is what’s left standing. Non-naturalism is also, he proclaims, the theory that answers best to human concerns. How could we know a nonnatural 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 for any such account is whether it could fit in with the best accounts we have of thinking and judgment. Our brains are upshots of Darwinian natural selection and other natural processes. No one well understands how this is, but the truth about reasons and how we know them must somehow fit in with a naturalistic story of our nature. How, then, could natural processes have equipped us to know nonnatural 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.”
Review of Parfit, On What Matters

Now true enough, not by genes alone are we caused to be what we are. Still, 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. A lot is mysterious in all this, no matter how “transparent and intelligible” we experience it all 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? Such questions plague any of us who know our brains don’t work by magic, and a satisfactory story must indicate how mathematical and normative judgments might operate. How can non-causal responses to non-natural properties figure in natural goings-on? I have some idea why intelligent, evolved organisms would get arithmetic and geometry right. Carpenters who bungle geometry mess up their work, and those who got such things right tended to reproduce more. Ethical judgment may help reproduction, but how could fitting the normative facts be part of the explanation? What helps reproduction, we might think, is eliciting cooperation on mutually advantageous terms. What does that have to do with nonnatural normative facts?

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 to a position that we haven’t learned how to reconcile with regarding human phenomena as natural—and with each decade we get further indications that naturalistically is how 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 general mode of approach to 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. That is Parfit’s method, in effect, whatever metanormative gloss on it might be best.

One seeming reason to distrust such inquiry is that different people who engage in it come to such opposed conclusions. A central aim in Parfit’s work is to undermine this qualm. “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”546i. Even Nietzsche would join in, Parfit argues in a fascinating chapter.

As for me, I am puzzled whether to accept the intuitions that ground Parfit’s arguments for moral convergence. I find the intuitions plausible but far from indubitable. It is remarkable, though, that Parfit is able to construct arguments for his conclusion that are plausible. Four tough, smart, and learned commentators don’t to my mind refute him. The lesson to draw isn’t
that Parfit has devised a clear demonstration that the three starting points he identifies will lead to the same peak in moral theory. Parfit himself claims no knock-down argument. Rather, the lesson is how very difficult it is to get plausible moral thinking to be systematic and yet depart from some sort of consequentialism. Attempts to accomplish this lead to the kinds of extravagances and special pleading that Parfit identifies in Scanlon and in Rawls.

Parfit’s doctrines and methods of argument do evoke Sidgwick, though the two massive treatises differ. Sidgwick establishes his intuitionism perfunctorily, but then over the four substantial books of the treatise makes it seem undeniable that the only form of moral theory that can account with any coherence for our moral intuitions is some form of consequentialism. Since his time, philosophers have raised possibilities that Sidgwick hadn’t noticed. Parfit’s survey of what could account for our intuitions as moral insights is by design less comprehensive than Sidgwick’s, but if it shows that contractualism in any plausible form leads to consequentialism, he has covered some of the main hopes of present day anti-consequentialists.

If there were no nonnatural 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 nonnatural. (To be sure, if nothing matters and there are no reasons, that can’t be reason to despair.) Even, though, if Parfit convinces us that mattering must be non-natural, when we ask “What if there’s no such thing as a non-natural property?” we need to consider two answers: that nothing matters, or that we were convinced wrongly. If no properties are nonnatural, either nothing matters or Parfit is wrong that mattering is nonnatural. Now it seems certain beyond doubt that things matter, whereas even if Parfit is right in his metatheory of mattering, we shouldn’t utterly trust that he is right. Suffering matters whatever the story of normative properties turns out to be.

For all else, it matters what matters. Whether mattering is nonnatural or not matters too, but not as much. Either way, it matters considerably if Parfit’s basic methods of discerning how things matter are reliable. I myself am queasy on whether they are, but I don’t see an alternative to his most basic methods. If, though, we question why these methods would work, nonnaturalism won’t tell us. I keep hoping that we can understand questions of what matters in a way that yields more insight.