

2014

Parfit's 'Triple Theory' and its Troubles

David McNaughton
Florida State University, dmcnaughton@fsu.edu

Piers Rawling
Florida Sate University, prawling@fsu.edu

Follow this and additional works at: http://digitalcommons.brockport.edu/phil_ex

 Part of the [Ethics and Political Philosophy Commons](#)

Recommended Citation

McNaughton, David and Rawling, Piers (2014) "Parfit's 'Triple Theory' and its Troubles," *Philosophic Exchange*: Vol. 44: No. 2, Article 3.

Available at: http://digitalcommons.brockport.edu/phil_ex/vol44/iss2/3

This Article is brought to you for free and open access by Digital Commons @Brockport. It has been accepted for inclusion in Philosophic Exchange by an authorized administrator of Digital Commons @Brockport. For more information, please contact kmyers@brockport.edu.



The College at
BROCKPORT
STATE UNIVERSITY OF NEW YORK

Parfit's 'Triple Theory' and its Troubles

David McNaughton
Florida State University

Piers Rawling
Florida State University

1. Derek Parfit's 'triple theory'

In his monumental and ambitious two volume work, *On What Matters*, Derek Parfit seeks to show that, suitably interpreted, three popular ethical theories—rule consequentialism, contractualism, and Kantianism—can be seen as not only compatible but mutually supporting. Parfit's book was originally entitled *Climbing the Mountain*; the three theories are like three climbers ascending the same mountain by different routes. As they emerge at the top they realize that they've all been heading for the same summit. Parfit regards this convergence as encouraging evidence that he is on the right track—if proponents of these differing theories actually turn out to agree with one another, doesn't this increase the likelihood that they are correct? We are skeptical.

First, we doubt that there is convergence—at least not if the three theories are as standardly portrayed. They have traditionally been viewed as rivals—not only espousing different methodologies, but also differing in their verdicts as to which acts are morally right. But Parfit, as we shall see, does not always adhere to standard portrayals. And there is an obvious problem with this aspect of his approach: the further he moves from the traditional portrayals, the weaker is his argument that the commonality between the theories is evidence of their (coincident) correctness—of course they agree if they've been modified to do so. Furthermore, even with his non-standard portrayals, there are still questions about whether Parfit achieves coincidence among the theories—his modifications may not go far enough.

Second, we doubt that any of these three theories are on the right track. All three are constructivist, as is Parfit's amalgam, and thus, in our view, share a common flaw. Constructivism about morality is the view that moral principles—rules about what is right and wrong—are 'constructed' in accord with some reliable and justifiable procedure that tests various proposed rules or principles, endorsing some and rejecting others. The moral status of an act is then metaphysically determined by the procedure via the principles it endorses. An act is morally wrong, for instance, because it is forbidden by such a principle, where this is not merely an epistemic

test—the fact that the act is forbidden by a principle endorsed by the procedure is what *makes* the act wrong.

These theories are constructivist to different degrees, however. Kantianism purports to be constructivist ‘all the way down’. But Parfit himself, and the contractualist, construct a moral superstructure on a foundation of reasons, which are not themselves constructed—in fact Parfit and the contractualist share to some extent our own view of reasons. And rule consequentialism may also be seen as constructing morality on a non-constructed foundation. Before exploring constructivism, then, it will be helpful to give our own non-constructivist view, and in the process we’ll look at yet another constructivist theory, reasons internalism, which is constructivist about reasons rather than morality (and which we also reject).

2. Our view

When we wonder what to do, we look for reasons that favor one course of action or another. These reasons to do something are standardly called practical reasons, to distinguish them from theoretical reasons, which are reasons for believing some proposition. On Donald Davidson’s influential conception, practical reasons are belief-desire pairs that both cause and rationalize our actions (Davidson 1963). Thus, suppose I fill the kettle with water, and turn on the power. What might explain this? Well, it may be that I want a cup of tea, and I believe that I need boiling water to prepare that beverage (a fact of which Americans seem unaware!). Then that desire and belief rationalize my action—they make it intelligible by showing how my action contributed to achieving my goal. In addition, according to Davidson, if I did indeed act for this reason (this belief-desire pair), then it caused me to act. I might have several belief-desire pairs that rationalize some act of mine, but what makes it the case that I acted for one pair rather than another is a matter, claims Davidson, of which pair caused me to act.

There is a large literature on Davidson’s view. But we may safely ignore it here, since we employ an alternative conception of reasons (what our colleague Al Mele calls, somewhat disparagingly, ‘reasons in the British sense’) according to which practical reasons are facts rather than psychological states. So, for example, the fact that it’s cold outside is a reason for you to wear a coat, regardless of whether you realize this. Crucially, in fact, there are two things

to be realized. Your reason is one: the fact that it's cold—call this F1. But there's also a second (F2): the fact that F1 is a reason for you to wear a coat. F1 is an uncontroversial kind of fact; but F2, if it is a fact, as we claim, is a *normative* fact. And this claim does give rise to controversy: some deny that there are any normative facts, but we won't defend this aspect of our position here. Beware confusing terminology, however. 'British' reasons are often called 'normative reasons' (in contrast to Davidsonian reasons, which are typically dubbed 'motivating reasons'), so that F1 (the fact that it's cold) is a normative reason for you to wear your coat. But F1 is *not* a normative fact. That honor goes to F2.

It is hard to explain what we mean by normative. But facts about what you have *reason* to do or believe, and facts about what you *ought* or *should* do or believe, are normative. Normative facts, then, are sometimes associated with standards that can fail to be met: if, for instance, you ought to do or believe something, and don't, you have failed to meet a standard, and are criticizable for this failure.

Note that whether you have a normative reason to do something is distinct from whether you desire to do it. Often, but not always, I desire to do things I will enjoy (and vice versa). But it is the fact that I will enjoy them that is the reason, not that I want to do them. (I might want to drink a can of paint, thinking I will enjoy it, but since I won't, I have no reason to.) Indeed, on our view, you might have a reason to do something even though you could not be brought to want to do it, even if, as explained below, you were fully informed about the non-normative facts, and reasoned in light of this information in an ideal way. Thus we are committed to the existence of what Bernard Williams calls 'external reasons'—we are 'reasons externalists'.¹

2.1. Internalism and externalism about reasons

Williams himself, on the other hand, espouses 'internalism' about reasons; you have a reason to do something only if, and because, you would (or, on another formulation, could) be motivated to do it under the assumption that you are fully informed (about non-normative matters) and 'procedurally rational'. On Williams' account, procedural rationality merely requires such things as

¹ See Williams 1981 and 1995.

adherence to the best deductive and inductive practices, and coherence of aims and intentions; crucially, it does not require having any particular prudential or moral concerns. Suppose you're in need of food. This is a reason for you to eat, according to reasons internalism, only if you would be motivated to eat if you were fully informed and procedurally rational. Thus, on Williams' view, if an anorexic cannot be brought to desire food, no matter how procedurally rational and (non-normatively) well informed he becomes, then he has no reason to eat.

As mentioned, reasons internalism is a constructivist theory, but only about reasons, at least in the first instance (there is a further debate about whether it also constructs right and wrong, but we won't go into that here). The reasons internalist's 'construction' starts with an agent's current 'motivational set', *S*, which includes not only her ordinary desires, but also 'such things as [her] dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects ... embodying [her] commitments' (Williams, 1981: 105). We then 'correct' her *S* so as to ensure that she is 'fully informed'—false non-normative beliefs are eliminated, and any missing relevant true non-normative beliefs are added.² And finally we set her to work following the rational procedures mentioned above—deduction, induction, and resolving conflicts among her aims and intentions. At the end of this process, what she has reason to do is exactly what she desires to do.

We reject this account. Whether or not an agent has a reason to act in a certain way does not depend on the outcome of some rational procedure. It may be that the anorexic has no reason to eat; but, if so, we contend (along with others—e.g., Parfit 1997), this is not dependent on rational procedure, information, and his current motivations in the way that reasons internalism supposes.

2.2. The 'two tier' conception of reasons

We hold, then, a two-tier view of practical reasons. At tier one are the reasons, which are characteristically non-normative facts, such as the fact that it's cold. At tier two are the facts that the tier one facts *are* reasons. Experience tells us that it's easy to muddle this distinction, so perhaps it helps to appreciate that the two tiers give

² Williams 1981:102-103; 1995: 36; see also Smith 1994: 156 and Parfit 1997: 100.

rise to two different possibilities of error: you might be mistaken about the weather (and believe it's warm, say); or you might fail to realize that cold weather is a reason to wear a coat.

It may help to get the distinction clear in one's mind to think of it in the following schematic manner.

Reasons (in the 'British sense')

Tier one fact (F1): it's cold outside.

Tier two fact (F2): F1 is a reason for you to wear your coat.

F1 is a non-normative fact.

F2 is a normative fact.

There are two possibilities of error:

Tier one error: you are mistaken about the weather.

Tier two error: you are mistaken about whether F1 is a reason to wear your coat.

2.3. 'Ought' and the weighing of reasons

On our view, practical reasoning consists in assessing what the reasons are, in some particular case, and how weighty they are. Some reasons are weightier than others and in many, but not all, cases the preponderance of weight will favor one particular course of action. So what ought I to do (ignoring for the sake of simplicity cases where the reasons do not favor a unique course of action)? An obvious suggestion would be: I ought to do what there is most reason to do. But there are problems with this suggestion. Suppose, for instance, a doctor has at her disposal three pills, about which she knows the following. The first would either cure the patient of his moderately debilitating illness, or kill him. If the first would kill him, the second would cure him; but if the first would cure him, the second would kill him. She has no way of finding out which one would kill and which cure. The third pill would merely provide partial relief.³ In this case (unbeknownst to the doctor, of course) the first pill would cure, so it's what the doctor has most reason to prescribe. But prescribing it is surely not what she ought to do.

³ This is a variant of one of Jackson's cases in Jackson 1991.

Given her state of ignorance, she ought to prescribe the third, even though she knows that this is not the act that she has most reason to perform—she knows that she has most reason to cure the patient, and so she knows that she either has most reason to prescribe the first pill or most reason to prescribe the second, but she knows that she can't find out which.

Or consider a more homely example of Prichard's (1932). I am approaching an unmarked traffic junction with limited line of sight. Do I have reason to slow down? Well, if no vehicle is approaching on the road I intend to cross then, no, I don't have a reason. But surely I *ought* to slow down—so that, counterintuitively perhaps, I ought to do something that I have no reason to do. Why? Because even if nothing is coming, it would be unreasonable to believe this without slowing down and checking, given my limited view. So what I ought to do depends not on facts that I couldn't reasonably be expected to know at the time for action, and not (as Prichard would have it) on my purely subjective assessment of the facts, but on the beliefs that are reasonable for me to have in my situation.

Our view resolves a puzzle about correctness of choices. Sometimes we want, in retrospect, to kick ourselves, and say we made the wrong choice, while at the same time acknowledging that our choice was the best we could make in our circumstances—how can this be? Suppose, for example, Anne has two routes to work, and the traffic is unpredictable, but her experience over the years is that route A is quicker on Mondays. So, today being Monday, she takes route A. But the traffic is awful. On arriving late to work she discovers from a co-worker who lives in the same neighborhood that route B was plain sailing. In one sense Anne made the wrong choice—she got stuck in traffic—but, in another sense, she made the correct choice: it was reasonable for her to expect the traffic to be better on route A. But she can't have made the incorrect choice *and* the correct choice in a univocal sense of correctness. We resolve this by proposing *two* standards of correctness here. What Anne had most reason to do was to take route B—taking route B would have been the correct choice in this sense. But she did as she ought, namely, take route A. And this was correct in the sense that it was based on what it was reasonable for her to believe in her situation.

Sometimes it is the case not only that I ought to do something, but that I *morally* ought to do it. What, then, is our account of moral obligation? Some reasons are moral reasons, and some are not. So, for example, that the pill would cure his disease is, normally, a

moral reason for prescribing it, while the fact that the chocolate would taste nice is a non-moral reason to eat it. (We are not claiming that the line between moral and non-moral reasons is a sharp one; no doubt there are grey areas.) Roughly speaking, to say that you have a moral obligation to A is to say that the only reasonable beliefs for you to hold in your circumstances are that there is most reason to A , and that the reasons to A are preponderantly moral.⁴ And to say that it would be morally wrong for you to A is say that it would be unreasonable for you to fail to believe that there is decisive moral reason against A -ing. But this isn't quite right, of course—consider the case of the pill prescribing doctor above. Morally she ought to prescribe the third pill, but it would not be reasonable for her to believe that this is what she has most reason to do. Matters, then (in both the moral and non-moral cases), are more complicated than we can go into here, involving, as they do, *degrees* of belief and decision theory.

But a crucial point for our purposes is that, in contrast to some of the views to come, the fact that an act would be morally wrong is certainly not a further independent reason against doing it. Suppose you are contemplating an act that would inflict pointless suffering. This is a decisive moral reason against it. But the fact that it would be unreasonable for you to fail to believe this is not a further independent reason against the act.

On our view, no more is needed for an ethical theory than the reasons, their strengths, and, when it comes to oughts, an appeal to what can be reasonably expected vis-à-vis an agent's epistemic powers. Constructivist theories, to which we now turn, have a more complicated structure.

3. Constructivism: the basic idea

⁴ There is an element of stipulation here. Imagine a case in which a reasonable person in your shoes would believe the following: there is most reason overall to A , with the reasons being preponderantly moral, yet there is most *moral* reason to B (although, of course, there is less reason *overall* to B than there is to A). We say that you're morally obligated to A , but others might reasonably deny this, and claim that you're morally obligated to B , although rationally you ought to A . (And there are also, of course, the decision theoretic complications mentioned below.)

In Plato's dialogue Socrates asks Euthyphro: are things pious because they are loved by the gods, or do the gods love them because they are pious? The Euthyphro question often serves to differentiate realist views about some field from constructivist ones. In many areas, the contending parties can agree on a biconditional. For example, we might all agree that there is a fruit in the basket if, and only if, were someone with normally functioning vision to look in the basket in good lighting conditions, she would have a perceptual experience as of fruit. However, a realist about physical objects holds that perceivers would have such an experience *because* there is fruit in the basket, whereas the phenomenalist, who is a species of constructivist about the external world, holds that there is fruit in the basket *because* perceivers would have such an experience. As it was sometimes put in the heyday of phenomenism, physical objects are constructed out of actual and possible sense-data.

The move from 'A because B' to 'B because A', we might call the 'Euthyphro switch', and it lies at the heart of the dispute that we shall be considering.

4. Kantian constructivism

Kantians famously offer a test that Kant (1993) dubbed the 'Categorical Imperative' (CI). If we want to know whether some action we are thinking of doing would be wrong we have, first, to work out what the principle behind our action is. So, for example, if I could really do with some money, I may be tempted to lie about my willingness or ability to repay you in order to get you to give me a loan. What is my personal principle (or maxim, as Kant calls it)? Well, perhaps it is this:

(LPM) I shall make lying promises whenever I can gain an advantage thereby.

I then ask myself whether I could will that the maxim on which I propose to act be a universal law. The details of how to apply this test—the CI test—are open to debate, but the general idea is that anything permissible for me is permissible for all. So I am committed to the general principle that it is morally permissible for everyone to make lying promises whenever they can gain an advantage thereby.

But, Kant claims, there is a contradiction in supposing that everybody could adopt, and thus live by, LPM. For if we try to imagine a world in which everyone adopts this principle, and everyone knows that everyone's doing this, then no-one would be able to gain an advantage by making a lying promise because no-one would trust anyone, and so no-one would believe any assurances given to her by others. Indeed, promising would be impossible in such a world. A promise is a form of words by which I bind myself to doing some particular action. As such, it depends on a convention, by which that form of words expresses an undertaking that others can rely on. But in a world where no-one is trustworthy, there could be no such convention, and the words would mean nothing. We can make this vivid by imagining an anthropologist reporting that, in the society she had been studying, the practice of promise-making exists but with this difference from ours: that no-one regarded anyone as bound by their promises. Such a report would be unintelligible because incoherent.

Note that Kant is *not* claiming that if everyone had license to tell lying promises, the results would be disastrous (although they might well be), but, rather, that it is incoherent to suppose that everyone could adopt LPM—LPM makes reference to promising, but its universal adoption would undercut the very possibility of making such binding commitments. Thus LPM fails the CI test—I cannot will it to be a universal law—and hence, claims Kant, it is morally impermissible to tell lying promises for personal gain.

As Kant quickly goes on to point out, the immoral person does not, of course, wish everyone to act like her. Quite the contrary. What she wishes is that *she* behave in this morally unacceptable way while others remain morally upright. Confidence tricksters require a world in which most are gullible and few are confidence tricksters. But, Kant urges, there can surely be no good reason why this immoral person, whose circumstances do not differ markedly from the situation of the rest of us, should be allowed to make an exception of herself.

Kantianism is clearly constructivist in character. It starts with a formal rational procedure: the CI test. If a maxim fails this test, it is unacceptable, and unacceptable because it fails the test, where the 'because' marks metaphysical dependence. The CI test is not merely an epistemic check—running a maxim through it is not analogous to checking your mental arithmetic by using a calculator—rather, failing the CI test is what *makes* a maxim unacceptable. Thus, for example, I am permitted to tell lying

promises for personal gain only if LPM passes the CI test; but it does not. And this failure, on the traditional Kantian approach, is what makes it morally wrong to tell lying promises for personal gain.

But there is another way of putting this—in terms of practical reasons. Is the fact that telling a lying promise would be to my personal advantage a (practical) reason for me to do so? Since LPM fails the CI test, the Kantian answer is ‘no’. And, as in the case of moral wrongness, personal gain fails to be a reason to tell lying promises *because* LPM fails the CI test, where the force of this ‘because’ is again not merely epistemic, but metaphysical—the CI test is not merely informing us when something isn’t a reason, rather, failing the CI test *makes* this the case. Thus the Kantian constructivist can be seen as not only constructivist about right and wrong, but also about practical reasons.

Kantians themselves, however, rarely express matters in terms of reasons—they typically focus instead upon rationality. And they claim that the whole domain of practical rationality (the rationality of action) is governed by the CI, with the latter’s authority deriving from its being, allegedly, constitutive of practical reasoning—anyone who reasons about what to do without applying the CI test is simply not engaged in practical reasoning. Indeed, one of the attractions of the Kantian test is that it presupposes nothing about what we have reason to do, other than the fact that practical reasoning is governed by the CI. It thus offers to circumscribe practical rationality simply by the application of a test that cannot be rejected without contradiction.

Kant’s theory has, however, been the object of a number of well-known criticisms. First, does the test rule out what we want it to rule out, and permit what we want it to permit? For instance, are there no circumstances under which personal gain is an overriding reason to tell a lying promise? What if the ‘personal gain’ is the preservation of your life, and this can only be accomplished by telling a lying promise to your kidnapper? We could perhaps try and have various different versions of LPM to account for the various cases (assuming we can work out what they all are), but even the following arguably fails the CI test:

(LPM*) I shall make lying promises whenever I can escape kidnappers thereby.

And, although telling such lying promises might be ineffective as a means of escape, it is surely not immoral. Or consider the following example, due to Parfit: the maxim 'give more to charity than the average' fails the standard version of the CI test, since clearly not everyone can do this at a given time. But surely, far from being morally unacceptable, this is a laudable goal.

Another issue is whether, assuming a Kantian verdict to be correct, the Kantian account of it is the right one. Is answering the question, 'why shouldn't I tell a lying promise in order to obtain money for some trivial pursuit?' really a matter of pointing out that some sort of contradiction or incoherence would ensue if LPM were to be universalized? And it is surely implausible to claim that what is wrong with murder, rape, and other heinous acts is that the maxims people follow in carrying them out yield contradictions (assuming that these maxims do yield such, which is doubtful).

Setting aside these objections, it is important to reiterate here that, for the Kantian, the verdict about whether some consideration is a reason to do something is metaphysically dependent upon a verdict issued by the CI test—it is the CI test that is metaphysically basic, not reasons. As Darwall (2006: 299) puts it, whether deliberation is "fully rational", in the Kantians' view, "is determined by internal, formal features of the deliberative process [roughly, the CI test], not by its responsiveness to independently establishable normative reasons". This is in marked opposition to our position. We hold that there *are* "independently establishable" reasons. And what makes an agent's deliberation "fully rational", or otherwise, is the extent to which it is appropriately responsive to these reasons (setting aside epistemic worries about the agent's ability to know the full details of her situation).

5. Partial constructivism

As we have just seen, Kantian constructivism can be seen as attempting to 'construct' not only right and wrong, but also practical reasons themselves, at least to some extent. But contractualism and Parfit's version of Kantianism both accept the existence of external reasons of the British sort, which are independent of any constructive process. And as far as we can see, rule consequentialism can also allow that there are reasons of this sort. These theories are, then, partially constructivist in the sense

that they ‘construct’ a superstructure on a non-constructed base. We’ll begin by looking at contractualism.

5.1. Contractualism

Contractualism has been endorsed by a number of writers, many of whom are influenced by Kant; we shall take Thomas Scanlon’s theory as our example.⁵ The central thought is that we should understand the correct moral code as the set of principles that ideally rational agents, all of whom are committed to living in mutual respect with each other, would agree to.

Whereas Kant asks of a principle whether rational agents could universally will it, Scanlon asks whether reasonable persons could reject it.

An act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behavior that no one could reasonably reject as a basis for informed, unforced general agreement. (Scanlon 1998: 153)

And, furthermore, when an act is wrong, this is *because* it would be disallowed by any such unrejectable set of principles—its being disallowed is what makes it wrong—hence the theory is constructivist about wrongness.

The aim is to ‘find principles [for the general regulation of behavior] that others who share this aim also could not reasonably reject’ (Scanlon 2002: 519), where someone may only object to some proposed principle if its general acceptance would place excessive or arbitrary burdens on her. Of course, compromise may be needed, since relieving a burden to one person may impose a burden on another. For example, someone might propose that if

⁵ Contractualism should not be confused with *contractarianism*. The latter theory thinks of morality as resulting from an actual or hypothetical contract between individuals in society for their mutual protection and advantage. Its main historical advocate and progenitor is Hobbes. Since actual individuals differ in the needs, vulnerabilities, and abilities that they bring to the bargaining table, one concern is that any contract that is acceptable to all parties may simply entrench existing power relations.

people dig holes in the sidewalk, they should place warning lights, but need do nothing else. However, what about the danger to the blind? Should we have lights that also emit warning noises? But now what about those who are both blind and deaf? What about the hopelessly inebriated, or the absent-minded philosopher such as Thales (who supposedly fell into a well while studying the heavens.) Should we have barriers? How sturdy do they need to be? The more difficult we make it for someone to fall into the hole, however, the more trouble the hole digger will have to go to. At some point *that* burden may become so great that she has a reasonable objection, so a balance must be struck.

Scanlon holds that there are British reasons independently of this procedure. That we have reason, for example, not to harm others, to help the distressed, and to treat people fairly is something that we bring to the construction. What work, then, does the construction do, on this theory? It establishes whether or not an act would be *morally wrong*. For Scanlon, the fact that an act would be wrong gives an agent a *further*, and decisive, reason not to do it, in addition to whatever reasons there *already are* against acting in that way. Scanlon's account of how he came to formulate this view neatly illustrates this point. He had, of course, always known that there are good reasons to feed the starving. But it was only, he reports, when he read Singer's 1972 that he realized that it was *wrong* not to contribute to famine relief organizations. He saw this as an extra and decisive reason to contribute. We, of course, would describe his awakening differently. It is not that its being wrong not to contribute is itself one of the reasons to contribute. Rather, what reading Singer might lead you to realize is that the moral reasons to contribute, even at some personal inconvenience, are so strong that they are overriding.

We note in passing that contractualism inherits from its Kantian origins a problem about the protection of animals and landscapes. For these are not rational agents, nor are they capable, even in theory, of coming to the table with objections to proposed principles. Scanlon, like Kant, recognizes that this is a problem, and tries to meet it by proposing a distinction between the central core of morality and a periphery. The central core concerns what we (persons) owe to each other. Creatures and things that are not persons can enter into the conversation, as it were, only if they have an advocate; someone who cares about their welfare or preservation sufficiently to make them important to that person. While this is better than Kant's appalling suggestion that all that is

wrong with torturing animals is that you might move on to people next (and that actually would be wrong!) it does not, we think, accord with the conviction of many that torturing animals for fun, say, would be a paradigm and central example of a moral wrong, on a par with doing the same to humans.

To sum up: according to contractualism, wrong acts are wrong because they are ruled out by principles that no one could reasonably reject. And the fact that an act is wrong is a further independent reason against performing it. One key question is whether, given Scanlon's commitment to British reasons, the constructivist superstructure does any work—when a principle is 'reasonably rejected' this is because reasons against it have won the day, so why can we not just appeal to these reasons directly, and decommission the superstructure?

5.2. Rule consequentialism

Rule consequentialism is the general type of a theory of which rule-utilitarianism is a particularly well-known species. Here, as with contractualism, the guiding thought is that, given the damage unconstrained people can do to each other, we need a set of rules for the regulation of our behavior. Since we want things to go as well as possible, the appropriate rules are those that, if generally accepted, would produce more good than any alternative set of rules.

An act is wrong if and only if it is forbidden by the code of rules whose internalization by the overwhelming majority ... has maximum expected value (Hooker 2000: 32).

Hooker's definition is framed as a biconditional, but rule consequentialism is constructivist since the dependence runs in one direction: its being forbidden by the best set of rules is what makes a wrong act wrong. But the foundation is not a construction—the good is metaphysically basic.

So rule consequentialism assesses rules, but not acts, in terms of their contribution to the good. Wrong actions are those that violate the rules. Thus an act may not be wrong and yet fail to produce the best ('right does not equal best') since the optimal rules must be, for example, simple enough to learn and sufficiently appealing that people will generally follow them. So they will often

lead us to do less good than we could. Rules that would be fine for angels might be disastrous for humans. And even if humans could be trained to follow them, the cost of inculcating them might be too high. As an example, at one point Hooker suggests that it would be better if people were impartially benevolent. However, so deeply ingrained is our tendency to care about, and therefore favor, friends and family, that it would be too costly to stamp out, so that the best set of rules for *us* must allow each of us to have 'special people' to whom we give preferential treatment.

The thought that the costs of societal change can be an important factor in determining the best set of rules leads to various objections; here are two. First, we doubt that the case for friendship rests upon the fact that it would be too costly to stamp it out. Second, a practice, however objectionable, might be permissible under the best set of rules because the costs of extirpating it would be too great. Imagine a society in which people so enjoy seeing bear baiting and dog-fighting that stamping it out would be prohibitively expensive in terms of time and resources. In such a sick society, the best set of rules would not forbid these barbaric practices and so, counterintuitively, they would not be wrong.

Although Hooker scarcely discusses reasons in the British sense, there is room in rule consequentialism for the thought that there are independent moral reasons for adopting it as the best moral theory. The idea that we should endorse practices that maximize the good is one such. In defending his view, Hooker also appeals to the idea that moral theories should be impartial as between persons—and this, again, is clearly offered as an independent moral ideal. However, rule consequentialism in its theory of how we should conduct ourselves on an act-by-act basis makes no reference to reasons—all that matters is that we follow the rules.

5.3. Parfit's Kantianism

As we have seen, Parfit holds that there are reasons in the British sense. Yet Parfit also believes that the Kantian is right to hold that an act can be right only if it can be rationally willed by all, but he gives a very different slant to this thought than does Kant or his followers.

Adapting Sidgwick's 1967 view, he holds that there are two different perspectives from which reasons can be discerned: the partial viewpoint, and the viewpoint of impartial benevolence. From the partial viewpoint, the act that will most benefit me or those to whom I am close, is the act that I have most reason to do (or to want to be done by others). From the perspective of impartial benevolence, the act that will produce the most overall good is the one I have most reason to do (or to want others to do). To look at matters from either of these viewpoints is rationally permissible. Here is how Parfit puts this view:

We ought ... to accept some *wide value-based objective* theory. On such views, when one of two possible choices would make things go in a way that would be impartially better, but some other choice would make things go better either for ourselves or for those to whom we have close ties, we often have sufficient reasons to make either choice. (2011 I: 186)

There are, then, both partial and impartial British reasons in Parfit's scheme. How do these relate to his version of Kantianism? Well, his 'Kantian Contractualist Formula' states that everyone ought to follow the set of principles whose universal acceptance everyone could rationally choose. And one might think that both partial and impartial reasons will play a role in the rational choice of moral principles. But this turns out not to be the case, as becomes evident when he argues from his Kantian formula to rule consequentialism. In outline, the argument runs as follows. Taking his Kantian formula as premiss, Parfit then notes that everyone could rationally choose whatever they would have sufficient reasons to choose. Next he considers principles that we have the strongest impartial reasons to choose—these are the principles whose universal acceptance would make things go best, the optimific principles. He then argues that these impartial reasons are not decisively outweighed by any conflicting reasons. Thus we have sufficient reasons to choose the optimific principles. Therefore, provided there are no other sets of principles that everyone has sufficient reasons to choose, the optimific principles are identical to the Kantian principles—the only principles that everyone has sufficient reasons to choose as the principles that everyone ought

to follow. And the claim that everyone ought to follow the optimific principles is rule consequentialism.

One key claim, then, is that:

- (E) No one's impartial reasons to choose [the optimific] principles would be decisively outweighed by any relevant conflicting reasons (I: 378)

As Parfit notes, 'we might have strong personal and partial reasons *not* to choose the optimific principles' (I: 379). So why mightn't these decisively outweigh the impartial reasons to choose the optimific principles? Parfit has his arguments (I: 379ff), but we won't go into them. Rather, our point here is that, in his view, the impartial reasons win out when choosing moral principles. But impartial, personal, and partial reasons can all play roles in choosing how to act on an individual occasion—or, at least, that would appear to be the upshot of the 'wide value-based objective theory'. There is, at the very least, a tension here. What if you are faced with a situation in which your reasons tell you to A and the rules tell you to B? Parfit might resolve this with his claim (which he shares with Scanlon) that the fact that an act would be wrong (i.e., violate the rules) is a further independent reason against it, so A-ing would have that mark against it, and thus you might yet have most reason to follow the rules. But this line suffers the problem that the reasons for which the rules were chosen have been counted twice in the weighing—once when they were weighed in choosing the rules, and again when the verdict of wrongness (to which they contributed via their role in rule choice) was weighed against A.

In addition to this difficulty, the selection of the optimific principles also makes it clear that, while retaining roughly the same form of words as the standard Kantian contractualist, Parfit relies on there being (impartial) British reasons to choose moral principles—reasons that are metaphysically basic, and not generated by any rational procedure. Parfit thus gives up the central plank of Kantianism—that there cannot be practical reasons external to the rational will. So any agreement between Parfit's Kantianism and contractualism or rule consequentialism is no evidence that a genuinely Kantian theory would agree with either.

6. A further problem for Parfit

A major bone of contention in recent ethics has been the existence of what are commonly called *constraints*. Of course, all moral considerations constrain action in some manner, but the term is here used in a somewhat technical sense. Are there acts that we should not do, even to bring about the greater good? If so, there are constraints against so acting. Many think, for example, that it is wrong to kill or torture the innocent even to obtain a great good—even, most poignantly, to prevent more innocent people being killed or tortured. Some conceive of constraints as absolute or exceptionless; others think of them as providing very strong reasons not to do certain things, but reasons that could be overridden in extreme circumstances. Call the latter threshold constraints. The Catholic philosopher, Elizabeth Anscombe, famously endorsed absolute constraints, as did Kant.

If anyone thinks, in advance, that it is open to question whether such an action as procuring the judicial execution of the innocent should be quite excluded from consideration—I do not want to argue with him; he shows a corrupt mind (Anscombe1958: 15).

Kant famously held that it was wrong to lie in any circumstances, even to save an innocent person from a would-be murderer. The justification of constraints is an integral part of the Kantian enterprise; it is what is involved in the second formulation of the Categorical Imperative: Never treat humanity, whether in your own person or in that of another, only as a means, but also at the same time as an end.

Parfit appears, at times, to endorse the notion that there are constraints. But of what kind? Not the absolute constraints beloved of Kant and Roman Catholic theology, but threshold constraints, perhaps. Now one advantage that is often claimed for rule consequentialism is that it follows common-sense morality in endorsing constraints. The best set of rules in a reasonably well-ordered society such as ours is likely to contain a rule against the deliberate killing of the innocent. (Since the most plausible version of rule consequentialism has a ‘disaster’ clause, to the effect that one can ignore the rules if disaster threatens as a result, there is no room in such a theory for absolute constraints.) But is this really a constraint? An advocate of constraints, whether absolute or threshold, holds that the fact that an act would involve, say, killing

an innocent, is *itself* a very strong reason against so acting. And it is a reason against in any circumstances and in any possible world in which there are morally accountable agents. Constraints, then, hold in all possible worlds. However, that is not how rule consequentialism sees the matter. For what rules are best depends on *empirical* issues, including how costly it would be to inculcate such a rule. And thus there could be worlds in which there was no rule corresponding to the constraint against killing the innocent. Imagine a world in which people are so cruel that it is very hard to eradicate their blood-lust (this is not, we fear, much of a stretch). A world in which, perhaps, there is something like gladiatorial contests, or hunting humans instead of foxes, or a version of the Hunger Games to appease the masses. It is not clear that in such a world the best set of rules would include a rule against killing the innocent, if it were too difficult to eradicate this unpleasant human trait. The 'constraints' that rule consequentialism endorses are thus not really constraints, but ersatz or quasi-such.

Insofar as Parfit endorses rule consequentialism as ordinarily understood, then, he must reject constraints in the Kantian sense.

To sum up the story so far, Kantians purport to generate British reasons. However, their construction seems shaky, at best. Contractualists allow that there are such reasons, but then build an unnecessary superstructure upon them. And rule consequentialists ignore them, at least when it comes to deciding what to do on an act-by-act basis. Parfit attempts to combine all three theories, along with British reasons, but his 'Triple Theory' cannot embody all their distinctive characteristics.

7. The 'legal model' of morality

Contractualism and rule consequentialism, in different ways, presuppose what we call a 'legal model' of morality. This is perhaps clearest in the case of rule consequentialism whose test seems well suited to answering the question of what would be the best code of laws. When legislating, it makes sense to insist that laws should be reasonably simple, be ones that people are not going to find too difficult to obey, etc. But why suppose that *morality* should be simple or easy to obey?

The contractualist, and Parfit, hold that the wrongness of an act is a further reason against performing it. This makes no sense, in our view, when we are considering moral wrongness, but is perfectly appropriate when it comes to what is *legally* forbidden.

There may, for example, be reasons for and against experimenting with some drug, but the fact that it is illegal to do so can, quite apart from issues of getting caught and being punished, be a further reason not to do it. That is because, in a reasonably well-ordered society, we have reason to obey the law. Parfit, the contractualist, and the rule consequentialist, all advocate, then, a legal model of morality.

This also has an impact on the pertinence of the age-old question ‘Why be moral?’ (i.e., why do what morality requires?) On our view, if construed as a request for reasons, this question is ill-formed. An act is morally forbidden just if (roughly) it would be reasonable for you to believe that there is a sufficient weight of (preponderantly moral) reasons against the act. So construed, the question becomes: what reason is there to refrain from doing what it is reasonably believed there is good reason not to do? This hardly makes sense. It does make sense, however, to ask ‘Why do what is legally required?’ And in certain circumstances, or in certain societies, we can intelligibly imagine that the answer would be that we do not have sufficient reason to do as the law bids, or, indeed, have no reason to do so, or even strong and overriding reasons against doing so. The inhabitant of Nazi Germany had overriding reason not to obey the iniquitous Nuremberg laws, and the starving person has overriding reason not to obey the law against stealing. However, at least two of the elements of the Triple Theory do invite the question ‘Why do what morality requires?’ Consider contractualism and rule consequentialism. Suppose the rules or principles that emerge from the process forbid actions of a certain type. I can surely ask myself whether I do have sufficient reason not to act in that way, and the answer will have to appeal to something other than the constructive process.

The Kantian may seem to avoid this difficulty, since rationality requires that we act as morality dictates, and the question: ‘What reason is there to act as rationality requires?’ is certainly one that neither has nor requires an answer. But the Kantian conception of rationality is far removed from the idea of what action is favored by the preponderance of British reasons, as can be seen if we spell out the Kantian conception of rationality. On this account, you have decisive reason not to kill, torture, etc. because your acting on your maxim will generate a contradiction in your will. Even if that is true, and even if that makes your act irrational, the Kantian answer, as we have seen, mislocates what is morally wrong with these acts—torture is not wrong because it (purportedly) suffers a

logical defect. So, as we see matters, it makes sense to ask on the Kantian view 'What reason do I have to do as morality (i.e., the CI test) requires?'

All three theories introduce simplicity where there is complexity, and are complex just where things are simple. There is a persistent tendency among philosophers to think that morality must, at least centrally, be about just one thing. That there must be some fundamental principle or procedure which, if properly applied, will both illuminate the nature of ethics and guide us through life's perplexities. We suspect that this is a philosopher's fantasy. Why suppose that there is some one thing that unites all moral issues other than their being moral? Why think, for example, that benevolence, fairness, and respect for autonomous choice have some common root beyond that of being morally relevant? And why suppose that the complexities of moral choice can be eliminated by the application of some formula or procedure? On the other hand, these constructivist accounts offer an over-complicated account of the theoretical structure of moral thinking. While they typically do not claim that ordinary people must understand or employ these complex procedures in order to know what is right or wrong, they do suggest that a complete understanding of moral enquiry is only available to someone who has mastered the ins and outs of a complex procedure. On our view, by contrast, all that is involved in moral thought is the weighing of reasons. The difficulty comes, not in understanding what moral thought is, but in the recognition and weighing of particular reasons in concrete cases. And no philosophical doctrine, as Aristotle pointed out long ago, is going to help us to do that well. In that sense, moral philosophy has rather less to tell us about how to live a virtuous life than these theories suppose.

Bibliography

- Anscombe, G. E. M. 1958 "Modern Moral Philosophy." *Philosophy* 33:1-19.
- Copp, D. 2006 (ed.) *The Oxford Handbook of Ethical Theory*. Oxford: Oxford University Press.
- Darwall, S. 2006. "Morality and Practical Reason: A Kantian Approach." In Copp 2006, 282-320.
- Davidson, D. 1963. "Actions, Reasons, and Causes." *The Journal of Philosophy* 60:685-700. Reprinted in Davidson's *Essays on Actions and Events*. Oxford: Clarendon Press, 1980.

- Hooker, B. 2000. *Ideal Code, Real World*. Oxford: Clarendon Press.
- Jackson, F. 1991. "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics* 101:461-482.
- Kant, I. 1993. *Grounding of the Metaphysics of Morals* (originally published 1785) trans. J. Ellington. Indianapolis: Hackett.
- Parfit, D. 1997. "Reasons and Motivation." *Proceedings of the Aristotelian Society*, supplementary volume 71: 99-130.
- Parfit, D. 2011. *On What Matters* (in two volumes), Oxford: Oxford University Press.
- Prichard, H.A. 1932. "Duty and Ignorance of Fact." Annual Philosophical Lecture, Henriette Hertz Trust, British Academy. Reprinted in Prichard, *Moral Obligation*. Oxford: Clarendon Press, 1949.
- Scanlon, T. M. 1998. *What We Owe To Each Other*. Cambridge, Mass.: Harvard University Press.
- Scanlon, T. M. 2002 "Reasons, Responsibility, and Reliance: Replies to Wallace, Dworkin, and Deigh." *Ethics* 112: 507-528.
- Sidgwick, H. 1967. *The Methods of Ethics*. London: Macmillan.
- Singer, P. 1972. "Famine, Affluence, and Morality." *Philosophy and Public Affairs* 1: 229-243.
- Smith, M. 1994. *The Moral Problem*. Oxford: Wiley-Blackwell.
- Williams, B. 1981. "Internal and External Reasons." In *Moral Luck*. Cambridge: Cambridge University Press.
- Williams, B. 1995. "Internal Reasons and the Obscurity of Blame." In *Making Sense of Humanity*. Cambridge: Cambridge University Press.