

Parfit on Reasons and Rule Consequentialism

Douglas W. Portmore

According to rule consequentialism, an act is morally permissible if and only if, and because, it is permitted by the ideal code of rules. And a given set of rules counts as the ideal code if and only if the expected value of the world in which it is adopted is greater than that of any alternative world in which some other set of rules is adopted. Now, I will argue that we don't always have sufficient reason to act as rule consequentialism requires us to act. And this means that either rule consequentialism is false or we don't always have sufficient reason to act as morality requires.¹ Either spells trouble for Parfit. On the one hand, if rule consequentialism is false, then Parfit's Triple Theory is false, for Triple Theory entails rule consequentialism. And although Parfit doesn't explicitly endorse Triple Theory, he claims that we have strong reasons to accept it.² On the other hand, if we don't always have sufficient reason to act as morality requires, then, by Parfit's own

¹ I'm not making the well-worn (and, I believe, false) claim that either rule consequentialism collapses into act consequentialism or it is incoherent. Rather, I'm claiming that we don't always have sufficient reason to act as rule consequentialism requires us to act. This is not a problem for those rule consequentialists who deny that we always have strong reasons to avoid acting wrongly. But it is a problem for Parfit, for he thinks that if we often have sufficient reason to act wrongly, then morality would be undermined (*OWM*, v. I, pp. 147–148).

² *OWM*, v. I, p. 418.

lights, morality is undermined, and his assumption that we always have strong reasons to avoid acting wrongly is called into question.³

The paper has the following structure. In sections 1, I argue that for us to have a consequence-based reason to perform an act, the reason for us to perform the act must be provided either by the fact that the act itself has good consequences or by the fact that the act is part of a set of acts that, if realized, would have good consequences. But, in section 2, I argue that we have a reason to perform an act in virtue of the fact that it is part of a set of acts that, if realized, would have good consequences only if we are able to see to it that this set and its good consequences are realized. Yet, as I show in section 3, rule consequentialism implies that we can be morally required to perform an act in virtue of this sort of fact even if we are unable to see to it that the relevant set and its good consequences are realized. Thus, from my arguments in section 1–3, it follows that, in many situations, we have no consequence-based reason for acting as rule consequentialism requires us to act. And, in section 4, I argue that whatever non-consequence-based reasons we may have for acting as rule consequentialism requires us to act in these situations are not strong enough to give us sufficient reason to so act. It follows, then, that either rule consequentialism is false or we do not always have

³ See *OWM*, v. I, pp. 147–148. Strictly speaking, Parfit claims not that morality would be undermined if we *sometimes* had decisive reason to act wrongly, but that it would be undermined if we *often* had decisive reason to act wrongly. Nevertheless, I'll spend most of the paper focused on the issue of whether we sometimes have decisive reason to act wrongly and leave a discussion of the frequency of these times and whether it's sufficient to undermine morality for the paper's conclusion.

sufficient reason to act as morality requires us to act. And, in section 5, I argue that this presents a dilemma for Parfit. Parfit must concede either that rule consequentialism (and, hence, Triple Theory) is false despite the putatively strong reasons that he believes we have for accepting it or that morality doesn't have the importance he seems to attribute to it given that its importance has been undermined by his own substantive account of morality.

1. Reasons and Consequences

It is relatively uncontroversial to suppose that if your ϕ -ing would have good consequences, then you have a reason to ϕ . For instance, if your pushing a button would have the effect of draining a mine shaft of otherwise rising floodwaters, thereby saving the five miners trapped inside, then you have at least some reason to push the button. But beyond supposing that the consequences of individual acts matter, we might further suppose that the consequences of sets of acts matter. For instance, we might suppose that you can have a reason to ϕ , not because your ϕ -ing would itself have good consequences, but because your ϕ -ing is part of a set of acts that, if realized, would have good consequences. Suppose, for instance, that the miners will be saved if and only if you push both button A at t_2 and button B at t_3 . Here, your pushing A at t_2 is not itself sufficient to do any good. But we might say that you have a reason to push A at t_2 , because it is part of a set of acts (i.e., the set consisting in your pushing both A at t_2 and B at t_3) that, if realized, would have good consequences.⁴

⁴ A consequence-based reason to ϕ is a reason to ϕ in virtue of the fact either that ϕ -ing would itself promote the impersonal good or that ϕ -ing is appropriately related to

Of course, one might object that, in determining whether you have a consequence-based reason to push A at t_2 , all that matters is whether the way the world would be if you were to push A at t_2 would be better than the way the world would be if you were to perform some alternative act instead.⁵ And it may be that if you were to push A at t_2 , you would not, as a matter of fact, follow up by pushing B at t_3 . In which case, the world would be no better for your pushing A at t_2 . According to this objection, then, you would have no consequence-based reason to push A at t_2 . For, on this view, what matters in determining whether you have a consequence-based reason to ϕ is not whether your ϕ -ing is part of a set of acts that, if realized, would have good consequences, but whether your ϕ -ing would itself have good consequence. Call this the *standard view*.

something else that would promote the impersonal good. This allows that there could potentially be both direct and indirect consequence-based reasons. Whereas the fact that S's ϕ -ing would promote the impersonal good would be a *direct* consequence-based reason for S to ϕ , the fact that S's ϕ -ing is appropriately related to an X—where X is something that, if realized, would promote the impersonal good—would be an *indirect* consequence-based reason for S to ϕ . So if there is a reason for S to push A at t_2 in virtue of the fact it is part of a set of acts (i.e., the set consisting in S's pushing both A at t_2 and B at t_3) that, if realized, would have good consequences, this would be an indirect consequence-based reason for S to push A at t_2 .

⁵ This amounts to the claim that the only consequence-based reasons are the direct ones.

The standard view is, I believe, mistaken.⁶ For there are cases in which it is plausible to suppose that you have a consequence-based reason to perform an act, not because your performing it would have good consequence, but because it is part of a set of acts that, if realized, would have good consequences. Here's one such case:

The Five: Five miners will be saved from drowning in rising floodwaters if and only if Faisal pushes both A at t_2 and B at t_3 . Now, as a matter of fact, Faisal *wouldn't* push B at t_3 even if he were to push A at t_2 . This is true, not because he *couldn't* push both buttons. Indeed, if he were to form at t_1 the intention to push both buttons, he would do precisely that. The reason, then, that he wouldn't push both buttons is that, as a matter of fact, he is not going to form at t_1 the intention to push both buttons—and this despite the fact that he has, as of the present (that is, as of t_0), the capacity to form this intention in response to the decisive reasons he has for doing so. Instead, he is, as a result of his malice towards the five miners, going to form at t_1 the intention to refrain from pushing B at t_3 irrespective of whether or not he has pushed A at t_2 . And, if this is the intention that he will form (and I'm stipulating that it is), then he *would* not push B at t_3 even if he were to push A at t_2 . Again, this is not to say that he *could* not see to it that he pushes both buttons. He could, for all he would need to do is to form at t_1 the intention to push both buttons, an intention that he would then carry out.

⁶ What follows is only a brief sketch of an argument against the standard view, but see Portmore 2015 for a more thorough argument.

Given both that Faisal would not push B at t_3 even if he were to push A at t_2 and that his pushing A at t_2 will do no good unless he follows up by pushing B at t_3 , it follows that, on the standard view, he has no reason (or, at least, no consequence-based reason) to push A at t_2 . For his pushing A at t_2 would have no good consequences. But, in this case, it seems plausible to suppose that he has a reason to push A at t_2 in virtue of the fact that it is part of a set of acts whose realization he can see to and whose realization would have good consequences: the saving of five lives.

Consider that it would be a mistake for Faisal to deliberate at t_0 as follows: “(1) I would not push B at t_3 even if I were to push A at t_2 . (2) There is no reason for me to push A at t_2 unless I’m going to follow up by pushing B at t_3 . (3) Given premise 1, I can conclude that I would not so follow up. (4) Therefore, I have no reason to push A at t_2 .” The problem with his deliberating in this way is that it inappropriately treats premise 1 as fixed for the purposes of his deliberations. But when deliberating about whether or not to do x , one should not hold fixed the fact that one is (or is not) going to do x , nor should one hold fixed the fact that one is (or is not) going to do y if whether one does y is just as much under one’s present control as whether one does x is.⁷ So when deliberating about whether or not to push A at t_2 , Faisal should not hold fixed the fact that he isn’t going to push A at t_2 , nor should he hold fixed the fact that he isn’t going to push B at t_3 .

Perhaps, though, the reader will balk at my suggestion that Faisal could presently see to it that he pushes both buttons even though he would not push B at t_3 even if he were to push A at t_2 . In that case, let me just stipulate that I’ll use the phrase ‘see to it’ such that:

⁷ See Portmore 2015 for a further defense of this claim.

S can at t see to it that a set of actions, α , will be realized if and only if there is some set of intentions such that S has at t the capacity to form these intentions at t' and α would (via some non-deviant causal process) be realized if S were to form these intentions at t' , where t' is immediately subsequent to t .

Still, some readers may question whether my being able to see to it (in this stipulative sense) that some set of acts and its good consequences are realized is itself *sufficient* for my having a reason to perform each of the acts in that set.⁸ But this does not matter for my purposes. For I will be arguing that my being able to see to it that some set of acts and its good consequences are realized is *necessary* for my having a consequence-based reason to perform the acts within that set that do not themselves have good consequences.

2. Two Types of Cases

To begin to see why we should think that this is a necessary condition, consider the following two types of cases:

Securability Cases: (1) S_1 is deliberating at t about whether or not to ϕ at t' ; (2) although S_1 's ϕ -ing at t' would not itself have good consequences, good consequences would ensue if both S_1 ϕ s at t' and S_2 ψ s at t'' ; (3) as a matter of fact, though, S_2 will not ψ at t'' ; but (4) S_1 *can* (but won't) at t see to it that S_2 ψ s at t'' (S_1 may or may not be identical to S_2 and $t < t' \leq t''$).

⁸ For more on this issue, see Portmore (2011), Ch. 6.

Insecurability Cases: (1) S_1 is deliberating at t about whether or not to ϕ at t' ; (2) although S_1 's ϕ -ing at t' would not itself have good consequences, good consequences would ensue if both S_1 ϕ s at t' and S_2 ψ s at t'' ; (3) as a matter of fact, though, S_2 will not ψ at t'' ; and (4) S_1 cannot at t see to it that S_2 ψ s at t'' (S_1 may or may not be identical to S_2 and $t < t' \leq t''$).

I'll argue that although S_1 has a reason to ϕ at t' in securability cases, this is not so in insecurability cases. And this, as I will show, suggests that my being able to see to it that some set of acts and its good consequences are realized is a necessary condition for my having a consequence-based reason to perform the acts within that set that do not themselves have good consequences.

For my purposes, it will be helpful to demonstrate how we can divide each of the above two types of cases into three main sub-types, depending both on whether S_1 is identical to S_2 and on whether t' is identical to t'' . Consider, for instance, the three main sub-types with respect to securability cases. First, there is the sub-type in which S_1 is identical to S_2 and t' is identical to t'' . One such case is *The Car*: while driving, Carl is deliberating about whether or not to accelerate as there is a slow-moving truck up ahead in his lane and a Ferrari coming up fast from behind in the only adjacent lane.⁹ Given the circumstances, his accelerating will have good consequences if and only if he does so while changing lanes. If he changes lanes without accelerating, he'll be rear-ended by the

⁹ I borrow this case, while modifying it for my own purposes, from Goldman (1978), p. 186.

Ferrari. And if he accelerates without changing lanes, he'll crash into the back of the truck. As a matter of fact, though, Carl would not change lanes even if he were to accelerate, because he is so frustrated with the slowing-moving truck that he intends to rear-end it. Yet it's extremely important that he instead accelerates while changing lanes, for that's the only way he'll make it to his meeting on time, and his job depends on it. Moreover, if he were to intend to accelerate while changing lanes, that's precisely what he would do.

Second, there is the sub-type in which S_1 is identical to S_2 but t' is earlier than t'' . *The Five* is such a case. Faisal's pushing A at t_2 will have good consequences if and only if he is going to follow up by pushing B at t_3 . But, unfortunately, Faisal would not push B at t_3 even if he were to push A at t_2 —although he certainly could see to it that he pushes both buttons merely by intending now to do so.

Third, there is the sub-type in which S_1 and S_2 are not identical and t' is earlier than t'' .¹⁰ To illustrate, consider a variation on *The Five*, which I'll call *The Second Five*: five miners will be saved if and only if Abe pushes A at t_2 and Beatrice pushes B at t_3 . Assume that although Beatrice would not push B at t_3 even if Abe were to push A at t_2 , this is only because Beatrice doesn't know that she needs to push B at t_3 , and, unfortunately, Abe has no intention of telling her this. Thus, even though Abe could see to it that the five are saved merely by pushing A at t_2 while simultaneously telling Beatrice at t_2 to push B at t_3 , Abe has no intention of doing either. In this case, Abe's pushing A at t_2 will have good consequences if and only if Beatrice is going to push B at

¹⁰ A fourth sub-type is where S_1 and S_2 are not identical but t' and t'' are. For reasons of space, I skip discussion of this sub-type.

t_3 , which is something that Beatrice isn't going to do given that Abe isn't going to tell her to do so.

It is fairly uncontroversial to suppose that, in *The Car*, Carl has a reason to accelerate in virtue of the fact that it is part of a set of acts (namely, the set consisting in his accelerating while changing lanes) that, if realized, would have good consequences: his keeping his job. The fact that Carl doesn't intend to change lanes doesn't mean that he has no reason to accelerate any more than the fact that, say, Hitler, facing an ultimatum from Britain, had no intention of ordering his troops to withdraw from Poland meant that he had no reason to do so. One can't avoid having a reason to ϕ merely by lacking the intention to ϕ , and this is so whether ϕ is a singular act (such as issuing an order) or a compound act (such as accelerating while changing lanes). Moreover, one can't avoid having a reason to perform each of the conjuncts of some compound act merely by lacking the intention to perform that compound act.

And why should it matter whether the individual acts that make up some compound act are contemporaneous or not if the agent has just as much present control over whether the compound act will be realized in either case? So if we claim that Carl has a reason to accelerate in *The Car* despite the fact that he doesn't intend to change lanes, then we should also claim that Faisal has a reason to push A at t_2 in *The Five* despite the fact that he doesn't intend to push B at t_3 . And likewise we should further claim that Abe has a reason to push A at t_2 in *The Second Five* despite the fact that he doesn't intend to tell Beatrice at t_2 to push B at t_3 . It seems that what's relevant is whether the agent can see to it that the relevant set of acts is realized, not whether the acts in the

set have identical agents or occur at identical times.¹¹ So we should, I think, say that S_1 has a reason to ϕ at t' in securability cases regardless of whether S_1 is identical to S_2 and regardless of whether t' is identical to t'' .

Should we go further and say that S_1 has a reason to ϕ at t' in insecurity cases—that is, regardless of whether S_1 can at t see to it that S_2 ψ s at t'' ? Should we say, for instance, that Carl would have a reason to accelerate even if he could not see to it that he changes lanes, as where, say, the steering wheel was stuck? And should we say that Abe has a reason to push A at t_2 even if he could not see to it that Beatrice pushes B at t_3 , as where, say, he has no way of communicating with Beatrice. It seems not, for such acts would be entirely pointless, if not harmful. It seems, then, that the mere fact that S_1 's ϕ -ing at t' is part of a set of acts (i.e., the set consisting in S_1 's ϕ -ing at t' and S_2 's ψ -ing at t'') that, if realized, would have good consequences is insufficient to provide S_1 with a reason at t to ϕ at t' . For such a fact to constitute a reason at t for S_1 to ϕ at t' , it must be that S_1 could at t see to it not only that she ϕ s at t' , but also that S_2 ψ s at t'' . We should conclude, therefore, that an agent has a reason to ϕ in virtue of the fact that her ϕ -ing is part of a set of acts that, if realized, would have good consequences only if she is able to see to it that this set and its good consequences are realized.

3. Rule Consequentialism

¹¹ Would it matter if, in *The Five*, Faisal were to undergo some psychology-preserving but identity-destroying process (such as, perhaps, fission or teletransportation) between t_2 and t_3 ? For more on this, see Portmore 2015.

I've argued that an agent has a reason to ϕ in virtue of the fact that her ϕ -ing is part of a set of acts that, if realized, would have good consequences only if she is able to see to it that this set and its good consequences are realized. Yet rule consequentialism implies that an agent can be morally required to ϕ in virtue of this sort of fact even if she is unable to see to it that the relevant set and its good consequences are realized. To see this, consider that, on rule consequentialism, whether I'm obliged to ϕ (e.g., to keep a promise) doesn't depend on whether my ϕ -ing would have better consequences than my not ϕ -ing, but on whether everyone's following a code of rules that requires me to ϕ would have better consequences than everyone's following any code of rules that doesn't require me to ϕ . Thus, on rule consequentialism, I'm obliged to ϕ because ϕ is a member of a set of acts that would have the best consequences: namely, the set consisting in everyone's following the ideal code. And rule consequentialism holds that I'm required to ϕ in virtue of its being a member of this set even if I cannot see to it that this set and its good consequences are realized. That is, I'm required to follow the ideal code even if I cannot see to it that everyone else follows the ideal code.

Now, strictly speaking, it is only one version of rule consequentialism that implies that I'm obliged to ϕ in virtue of the good consequences resulting from the set of acts consisting in everyone's following the ideal code—namely, the *universal following* version of *rule consequentialism*, or

UFRC: An act is wrong if and only if, and because, it is disallowed by the UF-optimific set of principles, where a set of rules, R1, is UF-optimific if and only if there is no alternative set of rules, R2, such that the expected value of the world in

which R2 is universally followed is greater than the expected value of the world in which R1 is universally followed.¹²

Shortly, I'll be considering other versions of rule consequentialism, but for now let me demonstrate that UFRC does indeed have the above-noted implication. To illustrate, consider

The Unsolved Climate Case: If everyone (or even just nearly everyone) makes the significant sacrifices required to drastically reduce their carbon footprints, then the climate disaster that will otherwise ensue in the next century will be averted. Unfortunately, no one is making, nor is anyone going to make, these sacrifices. Moreover, each individual has control over only his or her own actions. Thus, regardless of what any individual does, climate disaster is going to ensue. Indeed, let's assume that no individual can make the slightest difference to the likelihood that the disaster will occur, to when it will occur, or to how morally bad it will be if it occurs.¹³ But individuals can make a significant difference as to how well

¹² As Parfit admits, this needs to be revised, for there is the possibility that two or more sets of principles with incompatible prescriptions could be UF-optimific. See *OWM*, v. I, pp. 407–408.

¹³ Assume that there is no precise boundary between suffering more and not suffering more just as there is no precise boundary between being balding and not being balding. And assume that just as the addition or subtraction of single strand of hair from a man's head cannot make a difference as to whether or not he is balding, the addition or

their own families will fare over the next century by not pointlessly making the sacrifices required to drastically reduce their own carbon footprints.¹⁴

In *The Unsolved Climate Case*, UFRC implies (1) that I am required to make significant sacrifices so as to drastically reduce my own carbon footprint, (2) that I am required to do so in virtue of the fact that my doing so is part of a set of actions (namely, the set of actions consisting in everyone's following the UF-optimific set of rules) that, if realized, would have good consequences, and (3) that this is so despite the fact that I cannot see to it that this set or its good consequences are realized. The only way UFRC wouldn't have these implications is if the UF-optimific set of rules were such as to permit people in my situation to refrain from making such sacrifices. But since, in this case, everyone is in exactly the same situation that I'm in vis-à-vis potential climate disaster, any such rule would permit not just me, but everyone else, to refrain from making such

subtraction of a single person's efforts to reduce his or her carbon footprint (because the relevant particulates will be so widely dispersed over the Earth's atmosphere) cannot make a difference as to whether or not any person suffers more. Now one type of response to this sort of sorites paradox is to deny that there are imprecise and morally relevant boundaries of the sort that I'm presupposing. But this is not a response open to Parfit, for he allows for genuine imprecision with respect both to normative truths and to differences in value—see *OWM*, v. II, pp. 555–562.

¹⁴ I call this *The Unsolved Climate Case*, because it is an instance of what Parfit calls an *unsolved each-we dilemma*—see *OWM*, v. I, pp. 305–306.

sacrifices.¹⁵ And the UF-optimific set of principles would not permit everyone (or even nearly everyone) to refrain from making such sacrifices, for any such set of principles would be one that, if universally followed, would inevitably result in climate disaster, which is, I'll stipulate, much worse than a world in which everyone follows a set of rules requiring them to make the sacrifices needed to drastically reduce their carbon footprints.

It might be thought that the ideal (or optimific) code would include a rule saying that one is required to bear the burdens of doing one's part in some possible cooperative venture only if one's doing so would not be pointless due to the unwillingness of others to do their parts.¹⁶ But even if the ideal code would include such a rule in certain possible worlds, it would not include such a rule in the possible world that I'm imagining, which is one in which climate disaster would ensue if everyone (or nearly everyone) were to follow (or even accept) such a principle. After all, given how bad climate disaster is, it's

¹⁵ Even if the UF-optimific set of principles exempted a few thousand (perhaps, randomly selected) individuals from having to make significant sacrifices, UFRC would still require those who were not exempted to make significant sacrifices despite the fact that their doing so would do absolutely no good and be bad both for them and their families. And note that no set of principles that exempted more than just a few thousand individuals from having to make significant sacrifices would be UF-optimific, for I'm stipulating both that, if more than just a few thousand failed to make significant sacrifices, climate disaster would ensue and that no set of rules that permits us to act in ways that allows climate disaster to ensue could be UF-optimific given how bad climate disaster would be.

¹⁶ See, for instance, Hooker (2000), pp. 124-125.

clear that no code whose universal (or near universal) following (or acceptance) would result in climate disaster could possibly be the ideal code—that is, UF-optimific. So we know that the UFRC-optimific code of rules would not permit everyone (or even nearly everyone) to refrain from making significant sacrifices in *The Unsolved Climate Case*.

So UFRC implies that I am morally required to make significant sacrifices because my doing so is part of a set of actions that, if realized, would have good consequences, and UFRC implies this despite the fact that I cannot see to it that this set or its good consequences are realized, and despite the fact that my making these sacrifices would have no good consequences for anyone and some bad consequences for me and my family. Yet I've argued that the fact that some act is part of a set of acts that would, if realized, have good consequences constitutes a reason to perform that act only if the agent can see to it that the relevant set and its good consequences are realized. This means that unless there is some other fact that constitutes sufficient reason for me to make these sacrifices, UFRC must either be false or be such that agents don't always have sufficient reason to act as it requires. Either would, I believe, be disappointing for Parfit.

Of course, Parfit needn't endorse UFRC; he remains neutral as to which version of rule consequentialism is best. But, as I'll show, every version of rule consequentialism faces the same dilemma: either that version is false or morality isn't as significant as Parfit thinks. For, as I'll show, every version of rule consequentialism implies that agents are sometimes required to perform acts that they lack sufficient reason to perform, and if agents sometimes have decisive reason to do wrong, that would undermine morality's importance. So either the strong reasons that Parfit claims that we have for accepting rule

consequentialism are decisively opposed by the reasons that we have for rejecting it or morality isn't as important as Parfit seems to think.

Let me turn now to showing how all other versions of rule consequentialism face the same dilemma: the dilemma of being either false or relatively unimportant. All versions of rule consequentialism hold that an act is wrong if and only if, and because, it is disallowed by the ideal code—that is, the optimific set of rules. Different versions of rule consequentialism differ only with respect to how we are to determine which alternative code is ideal. The two main points of contention are (1) whether we are to look at the expected value of each alternative code's being *universally* or *partially* “adopted” and (2) whether we are to understand the “adoption” of a code in terms of *accepting* or *following* the code. On what I'll call *universal* versions of rule consequentialism (such as UFRC), we look only at the expected value of the various alternative codes being universally adopted. On *partial* versions of rule consequentialism, we look also, or instead, at the expected value of the various alternative codes being partially adopted. And whereas, on *following* versions of rule consequentialism, we look at the expected value of the various alternative codes being (partially and/or universally) followed, on *acceptance* versions of rule consequentialism, we look at the expected value of the various alternative codes being (partially and/or universally) accepted, where the acceptance of a code involves more than just a disposition to follow it. It involves, among other things, believing what the code says.

So there are three types of rule consequentialism besides UFRC: (1) universal acceptance rule consequentialism or UARC, (2) partial following rule consequentialism or PFRC, and (3) partial acceptance rule consequentialism or PARC. And, with these

distinctions in place, I'm now in position to explain how each of these other three face the same dilemma that UFRC faces, starting with PFRC.

PFRC is to be defined exactly as UFRC was defined above except that we are to replace 'partially' for 'universally' and 'PF' for 'UF' throughout. Now, there are many different versions of PFRC, for there are many different ways of understanding how we should think of the partial following of a code. On *fixed-rate* PFRC, for instance, we assess each alternative code by the expected value of its being followed by some fixed-rate—say, by 90% of the population. On *variable-rate* PFRC, we look at each alternative code and the expected value of its being followed at every rate from 0% to 100% (inclusive), and then take the average expected value of the consequences for all of these rates. The ideal code, then, is the one with the highest average. On Parfit's *best-at-every-rate* version of PFRC, "everyone ought to follow the rules whose being followed by any number of people rather than by no one would make things go [expectably] best."¹⁷

Unfortunately, each of these has significant problems. The problem with Parfit's best-at-every-rate version of PFRC is, as Michael Ridge points out, that it "entails that if there is no single code which is [expectably] best *for each and every single* [rate of following]...then nothing is morally required."¹⁸ In a couple of endnotes, Parfit makes clear that he is aware of this problem, and he seems to be inclined to adopt something like Ridge's variable-rate version in response to the worry.¹⁹ But, unfortunately, Ridge's variable-rate version of rule consequentialism is subject to its own devastating objections,

¹⁷ *OWM*, v. I, p. 319.

¹⁸ Ridge (2009), p. 68

¹⁹ *OWM*, v. I, pp. 469 and 479.

as are all the other versions of PFRC mentioned above.²⁰

The best version of PFRC is, perhaps, *maximizing-expectation-rate* PFRC, which holds that the ideal code is the one whose expected value is at least as high as the corresponding expected value of any alternative code.²¹ The *weighted* expected value of a code *at a given rate* of following is the product of the expected value of that code at that rate of following and the probability that that code will be followed at that rate. And the expected value of a code is just the sum of that code's weighted expected values at every rate of following from 0% to 100% (inclusive). Since I take this to be the best version of PFRC, I'll focus on it. But my arguments would work, *mutatis mutandis*, against any other version of PFRC.

If the expected value of a code is to be calculated in terms of subjective probabilities, then there is no way for maximizing-expectation-rate PFRC to avoid implying that agents will be required to make pointless sacrifices in certain more fully specified versions of *The Unsolved Climate Case*.²² After all, if the comparative value of a world in which climate disaster ensues is low enough and/or the subjective probability that nearly everyone will follow a code requiring significant sacrifices is high enough, then the ideal code—that is, the code with the highest expected value—will require agents to make significant sacrifices in *The Unsolved Climate Case* even though, as a

²⁰ See Tobia (2012).

²¹ Tobia (2012), p. 7.

²² Subjective probabilities depend (at least, in part) on the evidence available to the subject or subjects in question. By contrast, objective probabilities do not. If determinism is true, then the objective probability of any given event is either zero or one.

matter of fact, there is, in this case, no objective chance that anyone will be making these sacrifices.²³

Of course, the defender of maximizing-expectation-rate PFRC could claim that the expected value of a code is to be calculated in terms of objective probabilities. And, in that case, the ideal code would not require agents to make significant and pointless sacrifices in *The Unsolved Climate Case*. For one of the stipulations of the case is that “no one is making, nor is anyone going to make,” the necessary sacrifices. Thus, we can assume that there is zero objective chance that a code requiring significant sacrifices will be followed. In which case, the code with the highest expected value could not be a code requiring significant sacrifices. But even if maximizing-expectation-rate PFRC avoids problematic implications in *The Unsolved Climate Case*, it has problematic implications in the following case.

The Solved Climate Case: If everyone makes the fairly minor sacrifices required to drastically reduce their carbon footprints (assume that, in this possible world, the needed sacrifices are fairly minor), then the climate disaster that would

²³ Parfit may claim that he is giving an account of what agents ought to do in the subjective (evidence-relative) sense as opposed to the objective (fact-relative) sense. But the evidence available to a given agent may suggest that her failing to make significant sacrifices would make absolutely no difference as to whether climate disaster will ensue or to how bad it will be. And it is implausible to suppose that an agent would have a subjective obligation to make significant sacrifices if her evidence suggests that her doing so would make no positive difference.

otherwise ensue sometime in the next century will be averted. As it turns out, everyone is doing far more than their fair share, as this is what they believe that they ought to do. That is, everyone believes that they should do more than their fair share (making significant as opposed to minor sacrifices) so as to ensure that even if many don't do their fair share, climate disaster will nonetheless be averted. Yet, unbeknownst to anyone, everyone is indeed doing far more than his or her fair share. Thus, regardless of what any individual does, climate disaster is going to be averted. Indeed, no individual can make the slightest difference to the likelihood that the disaster will occur—it won't. But individuals can make a significant difference as to how well their own families will fare over the next century by not making any significant (as opposed to minor) sacrifices themselves.

In this case, there is no *objective* chance that anything other than a code requiring each to do more than his or her fair share will be followed (or accepted for that matter). So the ideal code will include such a requirement, for, on the objective version of maximizing-expectation-rate PFRC, any principle with a 100% objective chance of being adopted will necessarily be part of the ideal code—assuming, that is, that there is some code that includes that principle that has a non-zero weighted expected value at some rate of following. After all, the expected value of any code lacking that principle would be zero given that it will have a 100% chance of being followed at a zero rate and a 0% chance of being followed at any non-zero rate.

So, on the objective version of maximizing-expectation-rate PFRC, individuals

are, in *The Solved Climate Case*, required to make significant sacrifices even though these sacrifices would be entirely pointless and constitute doing far more than their fair shares. I find this implication of the objective version of maximizing-expectation-rate PFRC to be at least as problematic as the subjective version's implication in *The Unsolved Climate Case*. For, in neither case, does the fact that making these sacrifices is part of a set of acts that, if realized, would have good consequences seem to be any reason at all to make such sacrifices given that the relevant agents cannot see to it that this set of acts or its good consequences are realized. So I conclude that PFRC faces the same dilemma that UFRC faces.

What about UARC? On this view, obligations don't arise from the good consequences of a set of *actions* being *realized*, but rather from the good consequences of a set of *rules* being *accepted*. So, unlike UFRC and PFRC, UARC never requires an agent to ϕ in virtue of the fact that her ϕ -ing is part of a set of acts that, if realized, would have good consequences. But UARC does require an agent to ϕ in virtue of the fact that her ϕ -ing is required by a set of rules that, if universally accepted, would have good consequences. But suppose, as I've argued, that we shouldn't think that the fact that an agent's ϕ -ing is part of a set of acts that, if realized, would have good consequences constitutes a reason for her to ϕ if she can't see to it that this set or its good consequences are realized. In that case, we also shouldn't think that the fact that an agent's ϕ -ing is required by a set of rules that, if universally accepted, would have good consequences constitutes a reason for her to ϕ if she can't see to it that these rules are universally accepted.

So although I had earlier drawn the rather limited conclusion that an agent has a consequence-based reason to ϕ in virtue of the fact that her ϕ -ing is part of a set of acts that, if realized, would have good consequences only if she is able to see to it that this set and its good consequences are realized, it seems that I can now just as plausibly draw the more general conclusion that an agent has a consequence-based reason to ϕ in virtue of the fact that her ϕ -ing is appropriately related to X (where X is something that, if realized, would have good consequences) only if she is able to see to it that X and its good consequences are realized. Insofar as the good consequences of X matter with respect to an agent's ϕ -ing (whether X be everyone's accepting a set of rules or the set of acts consisting in everyone's following those rules), they matter only insofar as she can see to it that X's good consequences are realized.

It seems, then, that UARC has problematic implications in *The Unsolved Climate Case* just as UFRC did. UARC implies (1) that I am required to make significant sacrifices so as to drastically reduce my own carbon footprint, (2) that I am required to do so in virtue of the fact that my doing is required by the set of rules that, if universally accepted, would have good consequences, and (3) that this is so despite the fact that I cannot see to it that these rules are universally accepted or that the good consequences stemming from their universal acceptance are realized. So, like UFRC, UARC implies that I'm required to make pointless sacrifices in *The Unsolved Climate Case*. And just as PFRC has problematic implications in *The Solved Climate Case*, so does PARC. Both imply that I'm required to make significant sacrifices even though these sacrifices would be entirely pointless and constitute doing far more than my fair share. Thus, the move

from following versions to acceptance versions of rule consequentialism seems to be of no help in avoiding counterintuitive implications in the climate cases.

4. Other Reasons

I've argued that that the fact that an agent's ϕ -ing is appropriately related to some X that, if realized, would have good consequences counts as a reason for her to ϕ only if she can see to it that X and its good consequences are realized.²⁴ Yet, in certain types of situations, which I'll call *unrealizability situations*, rule consequentialism requires agents to ϕ in virtue of the fact that their ϕ -ing is appropriately related to some such X even though they cannot see to the realization of X or its good consequences. (Note that depending on which version of rule consequentialism we're considering at least one of the two climate cases will count as an unrealizability situation.) Of course, this doesn't mean that agents lack sufficient reason to act as rule consequentialism requires them to act in these situations. For even if the fact an agent's ϕ -ing is appropriately related to some such X doesn't itself count as a reason for her to ϕ , it may be that some other fact constitutes sufficient reason for her to ϕ . In this section, I explore this possibility and argue that there is no other fact that constitutes sufficient reason for agents to ϕ in unrealizability situations.

So, if the fact that an agent's ϕ -ing is appropriately related to some X that, if realized, would have good consequences is not a reason for her to ϕ in an unrealizability

²⁴ 'X' could, for instance, stand for 'the set of actions consisting in everyone's following the ideal code' or for 'the universal acceptance of a particular set of rules'. And the appropriate relation might be that of being a part of that set or being permitted by that set.

situation, then what other fact might constitute a reason for her to ϕ in that situation? One possibility is that some deontological consideration counts in favor of her ϕ -ing. But in many unrealizability situations there will be no deontological consideration that counts in favor of her ϕ -ing. In the two climate cases, for instance, there are no deontological considerations that count in favor of my making significant sacrifices.²⁵ After all, I won't be breaking any promises if I fail to make significant sacrifices. Moreover, I won't be harming anyone if I fail to make significant sacrifices, for it's stipulated in each case that my failing to make significant sacrifices will neither cause harm nor increase the risk of harm to anyone. And fairness doesn't speak in favor of my making significant sacrifices. Indeed, fairness speaks against requiring me to make significant sacrifices in the two climate cases. In *The Unsolved Climate Case*, it would be unfair for me to have to make significant sacrifices when no one else is making such sacrifices. And, in *The Solved Climate Case*, it would be unfair for me to have to make *significant* sacrifices when that constitutes doing far more than my fair share.²⁶ Lastly, it won't do to object: "What if everyone did that?" For, in *The Unsolved Climate Case*, it seems sufficient to reply:

²⁵ It is also unclear how a rule consequentialist could even appeal to deontological moral considerations, for the rule consequentialist holds that the only moral considerations are rule consequentialist ones.

²⁶ One might suggest what would be fair is my making the same sacrifices that everyone else is making. But I don't see why the fact that others are irrationally doing more than what each of us needs to be doing makes it unfair for me to do no more than what each of us needs to be doing. (And I'll just stipulate that, in *The Solved Climate Case*, others are irrationally doing more than what each us need to be doing.)

“Everyone *is* doing that.”²⁷ And, in *The Solved Climate Case*, it seems sufficient to reply: “There would be no problem if everyone made only minor sacrifices, for, in *The Solved Climate Case*, that’s all that’s needed to avert climate disaster.”

But even if there are no deontological considerations that count in favor of my making significant sacrifices in these cases, perhaps the mere fact that I’m morally required to make these sacrifices would itself count as a reason for me to do so. As it turns out, Parfit thinks that the fact that some act is morally required does constitute a reason (what he calls a *deontic reason*) to perform that act. I’m skeptical about this, but I must admit that, if Parfit’s right and if, as rule consequentialism supposes, I’m morally required to make these sacrifices, then there would be at least some reason for me to make these sacrifices. Perhaps, then, this deontic reason will (itself or in conjunction with other facts) constitute sufficient reason for me to make these sacrifices. But, as Parfit admits, the plausibility of supposing that we have sufficient reason to do what morality requires depends on what morality requires. He says, for instance, that if act consequentialism were true, then we could plausibly deny that we always have sufficient reason to do what’s morally required, for act consequentialism is very demanding. For instance, act consequentialism requires me to sacrifice my life if I could thereby increase

²⁷ Parfit seems to concede as much, see *OWM*, v. I, pp. 305–306 and 319. What we might ask instead is whether we could rationally will that everyone as opposed to no one makes significant sacrifices so as to avert climate disaster. But this question isn’t equivalent to asking whether any individual should make significant sacrifices, and, as I’ve argued, the answer to this question depends on whether the individual can see to it that sufficient others do their part.

the overall good by even the smallest of increments. And, as Parfit points out, it's implausible to suppose that we have sufficient reason to meet such stringent moral demands.²⁸ Parfit's thought seems to be that if morality turns out to be very demanding, then it will be implausible to suppose that we always have sufficient reason to act as morality requires. In which case, we should suppose that whatever deontic reason we have to do what's morally required would be decisively opposed by the reasons that we have not to comply with such extreme demands.

Yet it seems to me that rule consequentialism's demands are even more extreme and unreasonable than act consequentialism's demands. At least, whenever act consequentialism requires us to make significant sacrifices there is always some point to our making these sacrifices in that they would produce some greater good. But, as we've seen, rule consequentialism requires us to make significant sacrifices even when doing so is completely pointless, doing absolutely no good whatsoever. So even if the fact that some act is morally required is itself a reason to perform that act, it seems doubtful that it will be strong enough to provide us with sufficient reason to make the sorts of significant and unreasonable sacrifices that rule consequentialism requires us to make in the climate cases.

Although I'm quite skeptical about whether the fact that an act is morally required is itself a reason to perform it, I'm much less skeptical about whether the fact that an act is one's only blameless and/or justifiable (justifiable to others, that is) option is itself a reason to perform it. But it seems doubtful that such a fact could ever provide sufficient reason for our making the significant sacrifices that rule consequentialism requires us to

²⁸ *OWM*, v. I, pp. 148–149.

make in unrealizability cases. For it seems that whatever reasons we have to avoid performing blameworthy and/or unjustifiable actions will not oppose, but coincide with, the (other) reasons that we have to avoid making such sacrifices. And, of course, the reasons that we have to avoid performing blameworthy and/or unjustifiable actions can't possibly provide us with sufficient reason to make such sacrifices unless they oppose the reasons we have to avoid making these sacrifices.

Why think that whatever reasons we have to avoid performing blameworthy and/or unjustifiable actions will not oppose, but instead coincide with, the (other) reasons that we have to avoid making such sacrifices? Well, I've argued that apart from whatever reasons we have to avoid performing blameworthy or unjustifiable actions, we have decisive reason to avoid making significant sacrifices in the climate cases. After all, we cannot realize any good by making such sacrifices, and we can realize some significant good for ourselves and for our families by refraining from making such sacrifices. Moreover, there are, I've argued, no deontic or deontological reasons that countervail against these reasons that we have to avoid making such sacrifices. So the reasons that we have to avoid making such sacrifices are, I believe, decisive. And this, it seems, is sufficient to make our refraining from making such sacrifices both blameless and justifiable. For if we have decisive reason to do something, then others will have to concede that, insofar as they are rational, they would act as we would in the situation. And how can they blame us for acting as they would act insofar as they are rational?²⁹

Perhaps, there could be expressive reasons for making significant sacrifices in the climate cases. For it seems that, by making such sacrifices, one could express both one's

²⁹ For more on this, see Portmore (2011), Ch. 2.

concern for the environment and one's willingness to cooperate with others. And it may be that we have reason to express such sentiments even when no good consequences will come of it. But even if we do, it seems that these reasons will not be sufficiently strong to countervail the reasons we have to refrain from imposing significant hardships on ourselves and our families, especially when there may be other ways to express such sentiments that are both less costly and more effective. Let's suppose, then, that although there is no way to prevent climate disaster in *The Unsolved Climate Case*, there is a way to reduce the environmental problems resulting from landfills by doing one's part in the "Reduce, Reuse, Recycle" campaign. It would seem, then, that doing one's part in this campaign would be a less costly and more effective way of expressing both one's concern for the environment and one's willingness to cooperate with others in such ventures.

Lastly, one might claim that there is a reason not to be complicit in any wrongdoing and that, in failing to make significant sacrifices in *The Unsolved Climate Case*, I would be complicit in our collectively bringing about climate disaster, which is wrong. But I'm doubtful that my failing to make significant sacrifices in *The Unsolved Climate Case* would count as my being complicit in this collective wrongdoing when, as I'll stipulate, (1) my own actions made no difference at all, (2) I was willing, and stood ready, to make these significant sacrifices were there any signs from others that they were also so willing, and (3) I clearly expressed my willingness to participate in such a cooperative venture and even actively sought commitments from others to so participate but to no avail. But even if there were such a reason, it seems insufficient to countervail

the reason that I have not to pointlessly incur significant hardships for my family and myself.

Above, I've canvassed various sorts of reasons that we might have for making significant sacrifices in unrealizability cases: deontic reasons, expressive reasons, deontological reasons, consequence-based reasons, reasons to avoid being complicit in wrongdoing, and reasons to avoid performing blameworthy and/or unjustifiable actions. None of them, taken either individually or collectively, seem to provide sufficient reason for making such sacrifices. But might there not be some other fact that I've failed to consider that constitutes sufficient reason for me to make significant sacrifices in unrealizability cases? I doubt it. If neither the fact that makes my making such sacrifices morally required on rule consequentialism nor the fact that I'm morally required to make such sacrifices provide sufficient reason for me to make such sacrifices (and I've argued that neither does), then whether there would be sufficient reason for me to make such sacrifices would have to be a contingent matter, based on, say, my having promised to make such sacrifices. But presumably, for whatever the contingency we think of, there will always be a way of further specifying the climate cases such that the contingency is not met. For instance, we can just stipulate, as I have, that I have made no such promise. We should, therefore, conclude that we sometimes lack sufficient reason to act as rule consequentialism requires.

5. Conclusion

We've seen that rule consequentialism sometimes requires us to act in ways that we lack sufficient reason to act. This presents a dilemma for Parfit. Parfit must concede either that

rule consequentialism (and, hence, Triple Theory, which entails it) is false despite the putatively strong reasons that he believes we have for accepting it or that morality doesn't have the importance he seems to attribute to it given that it has been undermined by his own substantive account of morality. Parfit could respond, though, that morality would be undermined only if we *often* have decisive reason to act wrongly, and I've shown only that we *sometimes* have decisive reason to act wrongly. How often must these times be to undermine morality? Parfit never says. But it seems to me both that unrealizability cases (on the model of the two climate cases) arise all too frequently and that, in any case, morality would be significantly undermined in terms its putative rational authority even if we only sometimes have decisive reason to act wrongly. In any case, Parfit admits that "it would be bad if, in such cases [cases where moral duty and self-interest conflict], we and others would have sufficient reasons to act wrongly."³⁰

The above-discussed dilemma arises for Parfit, because he holds that "moral principles or theories are intended to answer questions about what *all* of us ought to do [emphasis in the original]."³¹ By contrast, principles and theories of rationality seem intended to answer a different kind of question: questions about what *each* of us ought to do. And it seems that it is the fact that Parfit thinks of moral principles in this way that leads him to argue for a moral theory that makes what an individual morally ought to do a function of the good consequences that would be realized if all of us acted in a certain way or accepted certain principles. But even if it's plausible to think of moral principles in this way (and I'm skeptical), it doesn't seem plausible to think of what an individual

³⁰ *OWM*, v. I, p. 143.

³¹ *OWM*, v. I, p. 306.

has most reason, all things considered, to do in this way.³²

Bibliography

Goldman, Holly S. [now H. M. Smith] (1978) 'Doing the Best One Can', in Alvin I.

Goldman and Jaegwon Kim (eds.) *Values and Morals* (Dordrecht: D. Reidel Publishing Company), pp. 185–214.

Hooker, B. (2000). *Ideal Code, Real World*. Oxford: Oxford University Press.

Parfit, Derek (2011) *On What Matters* (Oxford: Oxford University Press).

Portmore, D. W. (2015). "Acts, Attitudes, and Rational Control." Working draft available at <http://bit.ly/1Jetb8j>.

Portmore, D. W. (2013). "Perform Your Best Option," *Journal of Philosophy* 110: 436–459.

³² For helpful comments and discussions, I thank Justin D'Arms, Dale Dorsey, Pamela Hieronymi, Jeff Moriarty, Shyam Nair, Jussi Suikkanen, Simon Kirchin, two anonymous reviewers, and audiences at both the University of St. Andrews and the 2013 New Orleans Invitational Seminar in Ethics.

Portmore, Douglas W. (2011) *Commonsense Consequentialism: Wherein Morality Meets Rationality* (New York: Oxford University Press).

Ridge, Michael (2009) 'Climb Every Mountain?' *Ratio* 22, 59–77.

Tobia, Kevin (2012) 'Rule Consequentialism and the Problem of Partial Acceptance', *Ethical Theory and Moral Practice* (30 August), pp. 1-10, doi:10.1007/s10677-012-9382-3.