50

Rule-Consequentialism

Brad Hooker

1 Introduction

Just what is the connection between moral rightness and consequences? For nearly half a century now, consequentialists have divided themselves into different camps with respect to this question. Act-consequentialists believe that the moral rightness of an act depends entirely on whether the act's consequences are at least as good as that of any alternative act. Rule-consequentialists believe that the rightness of an act depends not on its own consequences, but rather on the consequences of a code of rules. […] This essay explores the prospects for rule-consequentialism.

2 What Constitutes Benefit?

Rule-consequentialism holds that any code of rules is to be evaluated in terms of how much *good* could reasonably be expected to result from the code. By 'good' here I mean whatever has non-instrumental value. What has non-instrumental value?

Utilitarians, who have been the most prominent kind of consequentialists, believe that the only thing with non-instrumental value is utility. All utilitarians have held that pleasure and the absence of pain are at least a large part of utility. Indeed, utilitarianism is often said to maintain that pleasure and the absence of pain are the *only* things that matter non-instrumentally. Certainly, this was the official view of the classic utilitarians Jeremy Bentham, J. S. Mill, and Henry Sidgwick – though in Sidgwick's case, equality seems to have independent weight as a tie breaker […]

Perhaps more common over the last thirty years has been the view that utility is constituted by the fulfillment of people's desires, even if these desires are for things other than pleasure. Many people, even when fully informed and thinking carefully, persistently want things in addition to pleasure. They care, for example, about knowing important truths, about achieving valuable goals, about having deep personal relationships, about living their lives in broad accordance with their own choices rather than always in accordance with someone else's […]. The pleasure these things can bring is of course important. Still, human beings can care about these things in themselves, i.e., in addition to whatever pleasure they bring.

This view, however, can be challenged. Some desires seem to be about things too unconnected with you for them to play a direct role in determining your good. Would your desiring that a stranger recovers fully from her illness make her recovery good for you, even if you never see or hear from her again? […] Naturally, the fulfillment of such a desire would *indirectly* benefit you *if* it brought you pleasure or peace of mind. But this is not to say that the fulfillment of your desire that the

Brad Hooker, "Rule-Consequentialism," from Hugh LaFollette, ed., *Blackwell Guide to Ethical Theory* (Blackwell Publishing, 2000), 183–204. Reprinted with permission of Wiley-Blackwell.

Ethical Theory: An Anthology, Second Edition. Edited by Russ Shafer-Landau.

stranger recovers herself constitutes a benefit to you. Rather if you get pleasure or peace of mind from the fulfillment of this desire, this pleasure or peace of mind constitutes a benefit to you (since you doubtless also desire pleasure and peace of mind for yourself).

The view that the fulfillment of your desires itself *constitutes* a benefit to you – if this view is to be at all plausible – will have to limit the desires in question. The only desires the fulfillment of which constitutes a benefit to you are your desires for states of affairs in which you are an essential constituent […], You are not an essential constituent of the state of affairs in which this stranger recovers. So her recovery doesn't itself constitute a benefit to you.

There seem to be reasons for further restrictions on the desires directly relevant to personal good. Think how bizarre desires can be. When we encounter particularly bizarre ones, we might begin to wonder whether the things are good simply because they are desired. Would my desiring to count all the blades of grass in the lawns on the street make this good for me? […] Whatever *pleasure* I get from the activity would be good for me. But it seems that the *desire-fulfillment as such* is worthless in this case. Intuitively, the fulfillment of my desires constitutes a benefit to me only if these desires are for the right things […] Indeed, some things seem to be desired because they are perceived as valuable, not valuable merely because desired or pleasant. […]

Views holding that something benefits a person if and only if it increases the person's pleasure or desire-fulfillment are in a sense "subjectivist" theories of personal good. For these theories make something's status as a benefit depend always on the person's subjective mental states. "Objectivist" theories claim that the contribution to personal good made by such things as important knowledge, important achievement, friendship, and autonomy is not exhausted by the extent to which these things bring people pleasure or fulfil their desires. These things can constitute benefits even when they don't increase pleasure. Likewise, they can constitute benefits even when they are not the objects of desire. Objectivist theories will typically add that pleasure is of course an objective good. These theories will also hold that ignorance, failure, friendlessness, servitude, and pain constitute harms.

For the most part, I will be neutral in this essay about which theory of personal good is best. *Usually* what gives people pleasure or enjoyment is also what satisfies their desires and involves the objective goods that could plausibly be listed. So usually we don't need to decide as among these theories of personal good.

But not always. Suppose the ruling elite believed that quantity of pleasure were all that matters. Then (to take a familiar leaf from *Nineteen Eighty-Four*) they might feel justified in manipulating the people and even giving them drugs that induce contentment but drain ambition and curiosity, if they thought such practices would maximize aggregate pleasure. Or suppose the ruling elite believed that the fulfillment of desire were all that matters. Again, the ruling elite might feel justified in manipulating the formation of preferences and development of desires such that these are easily satisfied. Now we can accept that – to some extent – our desires should be modified so that there is some reasonable hope of fulfilling them. But this could be pushed too far either in the name of maximizing pleasure or in the name of maximizing desire-fulfillment. A life could be maximally pleasurable, have maximum desire-fulfillment, and still be empty – if it lacked desires for friendship, achievement, knowledge, and autonomy.

3 Distribution

The term "rule-utilitarianism" is usually used to refer to theories that evaluate acts in terms of rules selected for their utility – i.e., for their effects on social well-being. The term "rule-consequentialism" is usually used to refer to a broader class of theories of which rule-utilitarian theories are a subclass. Rule-consequentialist theories evaluate acts in terms of rules selected for their good consequences. Non-utilitarian versions of rule-consequentialism say the consequences that matter are not limited to net effects on overall well-being. Most prominently, some versions of rule-consequentialism say that what matters are not only how much well-being results but also how it is distributed, in particular the fairness of alternative distributions. Table 50.1 might prove helpful.

Which version of rule-consequentialism is best? The problem with rule-utilitarianism is that it has the potential to be unfairly inegalitarian. […] Consider a set of rules which leaves each member of a smaller group very badly off, and each member of a much larger group very well off (Table 50.2).

Now if no alternative rule would provide greater net aggregate benefit, then utilitarians would endorse this

Table 50.1

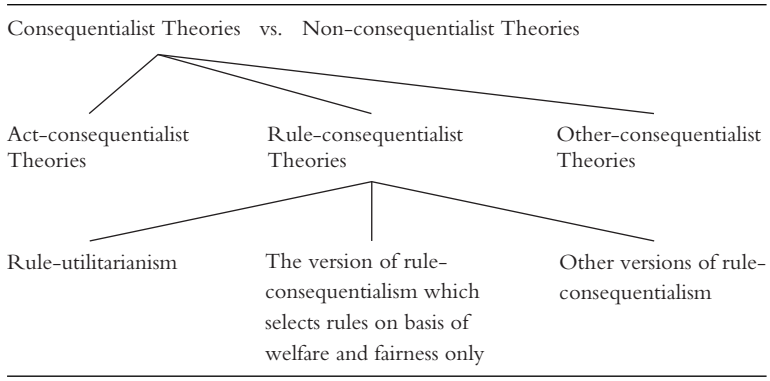

Table 50.2 Well-being

First Code:	*per person*	*per group*	*for both groups*
10,000 people in group A	1	10,000	
100,000 people in group B	10	1,000,000	
			1,010,000

Table 50.3 Well-being

Second Code:	*per person*	*per group*	*for both groups*
10,000 people in group A	8	80,000	
100,000 people in group B	9	900,000	
			980,000

code. Yet suppose the next best rule *from the point of view of utility* would be one with the results set out in Table 50.3.

Let us assume that the first code leaves the people in group A less well off for some reason other than that these people opted to work less hard or imprudently took bad risks. In that case, the second code seems morally superior to, because fairer than, the first code. This is why we should reject rule-utilitarianism in favor of a distribution-sensitive rule-consequentialism that considers fairness as well as well-being.

What are the relative weights given to well-being and fairness by this distribution-sensitive rule-consequentialism? Clearly, well-being does not have overriding weight. For there can be cases in which the amount of aggregate net benefit produced would not justify rules that were unfair to some group. That was what my schematic example above was meant to show.

Does fairness have overriding weight? This is particularly unsettled territory, since even what constitutes fairness is unclear. Nevertheless, we cannot rule out the possibility that some unfair practice so greatly increases overall well-being that the practice is justified. But it is certainly unclear where the threshold is for fairness to trump well-being. Perhaps the best we can say is that, in the choice between codes, judgment will be needed in balancing fairness against well-being. By evaluating rules in terms of two values (well-being and fairness) instead of one (well-being), distribution-sensitive rule-consequentialism is messier than rule-utilitarianism. Still, this seems to be a case where the more plausible theory is the messier one.

4 Criteria of Rightness versus Decision Procedures

Rule-consequentialism is often portrayed as merely part of a broader consequentialist theory. This broader

theory evaluates *all things* by their consequences. So it evaluates the desirability of acts by their consequences, the desirability of rules by their consequences, etc. The standard point to make along these lines is that, even if the rightness of an act depends on its consequences, better consequences will result if people do *not* try always to decide what to do by calculating consequences than if they try always to decide in this way. In other words, consequentialists can and should deny that

> On every occasion, an agent should decide which act to do by ascertaining which act has the greatest expected good.

Consequentialists agree that our *decision procedure* for day-to-day moral thinking should instead be as follows:

> At least normally, an agent should decide how to act by referring to tried and true rules, such as "Don't harm others", "Don't steal", "Keep your promises", "Tell the truth", etc.

Why? First, we frequently lack information about the probable consequences of various acts we might do. Where we cannot even estimate the consequences, we can hardly choose on the basis of maximizing the good. Second, we often do not have the time to collect this information. Third, human limitations and biases are such that we are not accurate calculators of the expected overall consequences of our alternatives. For example, most of us are biased in such a way that we tend to underestimate the harm to others of acts that would benefit us.

Now if there will be greater overall good where people are largely disposed to focus and act on non-consequentialist considerations, then consequentialism itself endorses such dispositions. So consequentialists advocate firm dispositions to follow certain rules, including firm dispositions not to harm others, not to steal, not to break promises, etc. Different consequentialists thus by and large agree about how people should do their day-to-day moral thinking.

What different kinds of consequentialists disagree about is what makes an act morally permissible, i.e., about the criterion for moral rightness.

> *Act-consequentialism* claims that an act is morally right (both permissible and required) if and only if the actual (or expected) good produced by *that particular act* would be at least as great as that of any other act open to the agent.

In contrast,

> *Rule-consequentialism* claims that an act is permissible if and only if it is allowed by a code that could reasonably be expected to result in as much good as could reasonably be expected to result from any other identifiable code.

The distinction between act-consequentialism's criterion of rightness and the dispositions it favors is important in many ways. It is important if we want to know what act-consequentialism wants from us. It is also important if act-consequentialism had better not conflict too sharply with our intuitive moral reactions. For if act-consequentialism claimed that we should always be focused on and motivated by calculations of what would maximize the good impartially conceived, many philosophers have thought it would be ridiculous. But the idea that act-consequentialism must make this ridiculous prescription is undermined by the distinction between act-consequentialism's criterion of rightness and the decision procedures it favors.

Nevertheless, the distinction is powerless to protect act-consequentialism from other objections. True, act-consequentialism's implications *about focus and motivation* are not as counter-intuitive as might initially be thought. But this is irrelevant to objections about act-consequentialism's criterion of rightness.

8 Arguments for Rule-Consequentialism

One argument for rule-consequentialism is that general internalization of rule-consequentialism would actually maximize the impartial good. The idea is that *from a*

purely consequentialist point of view rule-consequentialism seems better than act-consequentialism and all other theories.

Many act-consequentialists reply by invoking their distinction between their criterion of rightness and the decision procedure for day-to-day moral decisions. They admit act-consequentialism is not a good procedure for agents to use when deciding what to do. But they think this does not invalidate act-consequentialism's criterion of rightness. They would add that, even if rule-consequentialism is an optimal decision procedure, this would not entail that rule-consequentialism correctly identifies what makes right acts right and wrong acts wrong.

Let us turn, then, to arguments for rule-consequentialism other than the one that internalizing rule-consequentialism would maximize the good. Consider the moral code whose acceptance by society would be best, i.e., would maximize net good, impartially calculated. Shouldn't we try to follow that code? Isn't the code best for general adoption by the group of which we are members the one we should try to follow? These general thoughts about morality seem intuitively attractive and broadly rule-consequentialist.

And consider the related question "What if everyone felt free to do what you're doing?" This question may in the end prove to be an inadequate test of moral rightness. But there is no denying its initial appeal. And there is no denying that rule-consequentialism is an (at least initially) appealing interpretation of the test.

Rule-consequentialism thus taps into and develops familiar and intuitively plausible ideas about morality. Morality is to be understood as a social code, a collective enterprise, something people are to pursue together. And the elements of this code are to be evaluated in terms of both fairness and the overall effects on the well-being of individuals, impartially considered.

But rule-consequentialism's leading rivals all likewise emerge from attractive general ideas about morality. For example, act-utilitarianism can be seen as emerging from the idea that all that ultimately matters from the moral point of view is whether individuals are benefitted or harmed, that everything else is only instrumentally important. […] And act-consequentialism, the broader theory than act-utilitarianism, can be seen as emerging from the intuition that it can't be wrong to do what produces the most good. […]

Now consider moral contractualism, the theory that an act is right if and only if allowed by rules which could not be reasonably rejected by anyone motivated to find rules that no one with this same motivation could reject. Contractualism develops from the idea that morality consists of rules to which everyone would consent under appropriate conditions. This seems a very appealing general idea – moral rules grounded in reasonable agreement.

Consider yet another theory. The moral particularism of Jonathan Dancy (1993) builds on the idea that moral truth is found not in cold inflexible principles but rather through a finely tuned sensitivity to particular cases in all their rich complexity. Actually, to be distinct, moral particularism must go beyond the claim that there are some conflicts between competing moral considerations which are so difficult that agents would have to have fine moral sensitivity and judgement to resolve them correctly. To be distinct, moral particularism must be the view that what counts as a consideration at all can be decided only on a case by case basis. This is just how Dancy frames his theory: the very same consideration can count morally in favor of doing an action in one case, and against in another, and there are few if any considerations that must always count on the same side morally. (Dancy points out that such properties as moral rightness itself do *always* count morally in favor of an act.)

Finally, as I understand what has come to be called virtue ethics, this approach grows from the thought that right and wrong actions can be understood only in terms of choices that a fully virtuous person would make. This thought then suggests that we take the nature of and rationale for the virtues as the primary focal points for our moral philosophy.

Thus all these moral theories – rule-consequentialism, act-utilitarianism, act-consequentialism, contractualism, particularism, virtue ethics – tap into familiar and intuitively attractive general ideas about morality, though different ones. So no one could claim that any one theory is the only one with this feature. The conclusion to draw from this is simple. The fact that a theory arises from and develops attractive general ideas about morality is hardly enough to show that it is superior to all its rivals.

Now among the questions we can go on to ask about competing moral theories are (1) whether they are coherent and develop from initially attractive ideas about morality, and (2) whether the claims they end up making about right and wrong in various circumstances are intuitively plausible. I have already argued

that rule-consequentialism develops from attractive ideas about morality. But I shall not fully discuss here the objection that rule-consequentialism *incoherently* claims that maximizing the good is the overarching goal and then that following certain rules can be right even when breaking them would produce more good. I admit that if we start from an overarching commitment to maximize overall good, then our rule-consequentialism might be an incoherent account of moral rightness. But I propose our route to rule-consequentialism starts elsewhere: we don't start from, and indeed don't have, an overarching commitment to maximize overall good. If I am right about that, then this objection falls apart. […]

What other route to rule-consequentialism might there be? In the next few sections, I will show that rule-consequentialism's implications about what is right or wrong in particular circumstances match our confident moral convictions quite well. But let me immediately address the familiar challenge to the idea that moral theories are to be tested by their match with intuitions. The familiar challenge is that moral convictions are merely inherited prejudices and as such cannot provide good reason for anything.

In reply to this challenge, let me say I of course recognize that people from different cultures have different moral intuitions, as do people even from the same culture. We must always be willing to reconsider our moral intuitions. They are scarcely infallible.

But, while they are not infallible, they can be crucial. Suppose we have two moral theories which are each coherent developments of appealing general ideas about morality. Suppose one of these theories has implications that match our convictions quite closely, and the other has implications that conflict with many of our most confidently held moral convictions. In this case, I cannot see what could reasonably keep us from thinking better of the theory with the more intuitively plausible implications. Indeed, it seems to me that we are at least as confident about what is right in *some* specific kinds of situation as we are about any of the general ideas about morality that get developed into different moral theories such as Aristotelianism, Kantianism, contractualism, and act-consequentialism. This is why almost all moral philosophers are unable to resist "testing" these theories by comparing the judgments that follow from them with our confident convictions about right and wrong in various kinds of situations.

Let me take stock. I've suggested three different ways of arguing for rule-consequentialism.

One is that rule-consequentialism is, from a purely consequentialist point of view, best. I myself am not relying on this argument.

The second is that rule-consequentialism develops from some very attractive general ideas about morality. Though this is an important feature of rule-consequentialism, I acknowledge rule-consequentialism is hardly the only theory that plugs into or develops from attractive general ideas about morality. So the fact that a theory is a coherent development of some initially very attractive ideas is not enough to make it superior to all its rivals.

The third argument for rule-consequentialism is that we can reach a reflective equilibrium between rule-consequentialism and our confident moral convictions. At least *some* moral convictions seem more secure than any theory that could oppose them. If this is right, then appeal to reflective equilibrium between abstract theory and moral conviction must be part of the defense of rule-consequentialism.

9 Rule-Consequentialism on Prohibitions

Whatever act-consequentialism says about day-to-day moral thinking, act-consequentialism's criterion of moral rightness entails that *whenever* killing an innocent person, or stealing, or breaking a promise, etc., would maximize the good, such acts would be morally right. W. D. Ross put forward the following example (Table 50.4) to illustrate that keeping one's promises

Table 50.4 Numbers below represent units of good

	Effect on person A	*Effect on person B*	*Total good*
Keeping promise to A	1000	0	1000
Keeping promise to A	0	1001	1001

can be right even when this would produce *slightly* less good (Ross 1930: 34–5):

Most of us would agree with Ross that keeping the promise would be morally right in this case. Act-consequentialism, of course, favors breaking the promise in this case, since that is the alternative with the most good. So, if we agree with Ross about this case, we must reject act-consequentialism.

Most of us also believe (as Ross went on to observe) that, if breaking the promise would produce *much greater* good than keeping it, breaking the promise could be right. We believe parallel things about inflicting harm on innocent people, stealing, lying, etc. Thus most of us reject what is sometimes called "absolutism" in ethics. Absolutists hold that certain acts (e.g., physical attack on the innocent, promise-breaking, stealing, lying) are *always* wrong, even when they would prevent the most extreme *disasters*.

Absolutism and act-consequentialism are, we might say, two ends of a spectrum. Whereas absolutism never permits certain kinds of act, even when necessary to prevent extreme disaster, act-consequentialism insists such act are right not only when a great disaster is at stake but also when a *marginal* gain in net good is in the offing. Act-consequentialists seem mistaken about these cases of marginal gain, just as absolutists seem mistaken about the disaster cases. Thus, absolutism seems to go too far in one direction, act-consequentialism in the other.

Rule-consequentialism, on the other hand, concurs with our beliefs both about when we can, and when we cannot, do normally forbidden acts for the sake of the overall good. It claims that individual acts of murder, torture, promise-breaking, and so on, can be wrong even when they result in somewhat more good than not doing them would. The rule-consequentialist reason for this is that the general internalization of a code prohibiting murder, torture, promise-breaking, and so on would clearly result in more good than general internalization of a code with no prohibitions on such acts.

Another rule whose general internalization would be optimal is a rule telling us to do what is necessary to prevent disasters. This rule is relevant when the only way to prevent a disaster is to break a promise or do some other normally prohibited act. In such cases, rule-consequentialism holds that the normally prohibited act should be done. I mention this rule about preventing disaster because its existence undermines the objection that rule-consequentialism would, in a counterintuitive way, prescribe sticking to rules even when this would result in disaster.

10 Doing Good for Others

Morality paradigmatically requires us to be willing to make sacrifices for others. Yet act-consequentialism is widely accused of going too far here too. Utility, impartially calculated, would be maximized if I gave away most of my material goods to the appropriate charities. Giving away most of my material goods is therefore required of me by (most versions of) act-consequentialism. I should probably even change to some more lucrative employment so that I would then have more money to give to charity. […] I could make much more money as a corporate lawyer, banker, stockbroker, accountant, gossip-columnist, or bounty-hunter than as an employee of a philosophy department. If people should be willing to make any sacrifices that are smaller than the benefits thereby secured for others, then I should move to the better paying job so that I will have a bigger salary to contribute to the needy. With a bigger salary, I would then have to give an even larger percentage of my earnings to aid agencies. The result would be a life of devoted money-making – only then to deny myself virtually all the rewards I could buy for myself with the money. After all, from an act-consequentialist perspective, my own enjoyment is insignificant compared to the very lives of those who would be saved by my additional contributions. Such reflections give special poignancy to Shelly Kagan's remark: "Given the parameters of the actual world, there is no question that [maximally] promoting the good would require a life of hardship, self-denial, and austerity" (1989: 360).

But many of us may on reflection think that it would be *morally unreasonable* to demand this level of self-sacrifice for the sake of others[1] However praiseworthy such self-sacrifice may be, most of us are quite confident that perpetual self-impoverishment for the sake of strangers is above and beyond what morality *requires* of us.

I have been discussing the objection that act-consequentialism requires us to make *huge* sacrifices in order to maximize our contribution to famine relief. Act-consequentialism also requires self-sacrifice even when the benefit to the other person is only *slightly*

larger than the cost to the agent. Consider, for example, the corner office in our building. Offices are allotted on the basis of seniority. Suppose you are the most senior person who might want this corner office. But if you do not take it, it will go to an acquaintance who spends ten percent more time in her office than you do in yours. Suppose we therefore reasonably guess that she would benefit a bit more from moving into this office than you would. This is not a life and death matter. Nor will she be so depressed by not getting the corner office that her work or domestic life will be seriously compromised. Nevertheless, she would get a bit more enjoyment out of the better office than you would. But you still take it for yourself. No one would think you unreasonable or immoral for doing so. Except in special circumstances, morality does not, we think, really *require* you to sacrifice your own good for the sake of slightly larger gains to others.

I have offered two objections about the demands of act-consequentialism. (1) Act-consequentialism requires *huge* sacrifices from you. (2) Act-consequentialism requires you to sacrifice your own good even when the aggregate good will be only *slightly* increased by the your sacrifice. In both ways, act-consequentialism is *unreasonably demanding*.

In contrast, rule-consequentialism would *not* require you to pass up the corner office and let your colleague have it. You are certainly permitted to do that if you want, but rule-consequentialism would not *require* such impartiality in your decisions about what to do with your own time, energy, money, or place in line. The rules the internalization of which could reasonably be thought to produce the most good would *allow* each person considerable partiality towards self (and even *require* partiality towards friends and family. […] For, as I noted earlier, the costs of getting a complete impartiality internalized by each new generation would be prohibitive.

Likewise, whereas act-consequentialism requires huge sacrifices for the sake of maximizing the good, rule-consequentialism seems not to require more than a reasonable amount of sacrifice for this purpose. Why? A rule-consequentialist might point out that, if everyone relatively well off in the world were to contribute quite modest amounts to the best aid agencies, the worst elements of poverty could be overcome.

The World Bank has been calling for contributions from the rich countries of 0.7 percent of GDP, the current average being less than half that. Much of this aid does not go to the most needy, but instead to countries that offer business for, or military alliances with, the donor country. The UN estimates that if merely 60 percent of the aid that the rich countries now give (i.e., 60 percent of about $57 billion) were intelligently spent on providing basic health services and clean water and on eliminating illiteracy, these problems could be fixed (*The Economist*, June 22, 1996: 64).

A rule-consequentialist will be interested in redistribution beyond what is required to secure the very basic necessities. But even after including these other potential benefits in the cost-benefit analysis, we might well conclude that the amount the world's relatively well off would each be required to give would not be unreasonably severe. […]

Consider the following example. Walking along a deserted road on your way to the airport for a flight to the other side of the world, you see a child drowning in a shallow pool beside the road. You could easily save the child, at no risk to yourself. But if you do save the child, you will miss your flight and lose the cost of the nonrefundable ticket.[2]

Everyone agrees you are obligated to save the child. This is true even if you are not terribly rich. Suppose the ticket costs as much as a tenth of your annual income. You would still be morally wrong not to make the sacrifice and save the child. And even if the probability of the child's drowning without your rescue is less than 100 percent – suppose, for example, it is 80 percent – you are obligated to sacrifice your ticket to save the child.

Now consider a variant of the example. […] You and I are walking to the airport when we see two small children drowning in a lake. You and I could each easily save the children, at no risk to ourselves. The two children are positioned in the lake in such a way that you and I could each save one and still get to our flights. But if one of us saves both children, he will miss his flight. Suppose you save one child, but I do nothing. Surely, you should now save the other.

Yet, were rule-consequentialism framed in terms of 100 percent compliance, how could it tell you to save the other? With 100 percent compliance, there would be no need for you to save the second child. With 100 percent compliance, once you'd done your share, you'd have done all that was needed. The rule that would be best given 100 percent compliance would presumably not require you to sacrifice more than you would have to sacrifice if everyone did their part. But if this rule is

applied to our case, where I am in fact not coming to the rescue, you are *not* obligated to save this child. This is clearly an implausible implication.

But I argued that rule-consequentialism should be framed in terms of less that 100 percent compliance. If rule-consequentialism is framed in terms of 90 percent compliance, we can envisage that there is a need for rules about how to act when others around you aren't doing their part. The rule might be, "When you happen to be surrounded by others who are not helping, then prevent disaster even if this involves doing more than you would have to do if the others helped." This rule *would* require you to save the second child from the shallow pond.

But if the world we live in – the real world – is one where partial compliance is ubiquitous, then a rule requiring you to make up for the non-compliance of others could become unreasonably demanding. Just how much would rule-consequentialism require you to make up for non-compliance by other people in a position to help? In earlier work, I assumed that rule-consequentialism would formulate a rule about aiding the needy in terms of a *fairly precise* level of contribution or sacrifice to the reduction of world poverty.

I now think this approach is hopeless. Consider a concrete moral code that could reasonably be expected to produce at least as much good as any other we can identify. It would contain rules requiring us not to injure others physically, not to steal, not to break promises, not to lie, etc. These rules might have *some* exceptions built into them (though not a general break-the-rule-whenever-you-could-thereby-produce-more-good exception, nor an unlimited set of much more specific exceptions). Nonetheless, there is pressure to have *fairly general* rules that can be applied to a wide array of situations. Oxfam's petitioning the rich to help the very poor is hardly the only situation where some people have an opportunity to help others at relatively little cost to themselves. There will be situations where the rich can help other rich, situations where poor can help other poor, even situations where the poor can help the rich. And there will be situations where the help needs to be in the form of physical effort, other situations where the help needs to be in the form of money or time.

Given all this, perhaps the optimific rule for such a world would not be "the rich should give the very needy at least precisely n percent of their annual income", but rather "people should help others in great need when they can do so at modest to themselves, cost being assessed aggregatively, not iteratively" (Cullity 1995: 293–5). Such a rule would apply in a wide array of situations – indeed, whenever some person can help another in great need. It is not limited to what the rich should do nor to what should be done concerning world poverty.

But because cost to the agent is to be assessed aggregatively rather than iteratively, the rule does not require one to help another in great need whenever the cost of helping *on that particular occasion* is modest. Having to help others whenever doing so *on that occasion* involves modest cost could easily be very costly. For each of us faces an indefinitely long string of such occasions, because any day on which we could give money to UNICEF or Oxfam counts as such an occasion. But many small sacrifices added together can amount to a huge sacrifice. The end of that road is self-impoverishment. If I am right, rule-consequentialism instead endorses a rule requiring sacrifices over the course of your life that add up to something significant. It allows but does not require personal sacrifice beyond this point.

I propose that this rule *would* have good consequences even in possible worlds that are either much poorer or much richer than ours. I don't have space here to argue either that rule-consequentialism would indeed end up with this rule in *all* possible worlds, or that this rule *always* has intuitively acceptable consequences. I mention this rule only in order to sketch one way in which a defense of rule-consequentialism might go. […]

11 Conclusion

Rule-consequentialism has an uncertain future. It needs to be carefully formulated if it is to avoid being a sitting target. In this essay, I have tried to improve its defenses by fine-tuning its formulation. I have also argued here that the theory develops from appealing general beliefs about morality, that it does not collapse into act-consequentialism, and that it coheres well with our intuitions about moral prohibitions and permissible partiality. As I see things, the theory is healthy now. But it is hardly invulnerable. Like someone walking through a dangerous city who has so far managed to fight off muggers emerging from behind every corner, the theory might meet an attack it cannot survive. I am curious to see whether that happens.