University of Reading

The role(s) of rules in consequentialist ethics

Book or Report Section

Accepted Version

Hooker, B. (2020) The role(s) of rules in consequentialist ethics. In: Portmore, D. (ed.) Oxford Handbook of Consequentialism. Oxford University Press, New York. Available at https://centaur.reading.ac.uk/89644/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

Publisher: Oxford University Press

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

The Role(s) of Rules in Consequentialist Ethics

Brad Hooker

Abstract

After preliminaries concerning different accounts of the good and the distinction between actual-consequence consequentialism and expected-value consequentialism, this paper explains why consequentialists should prescribe a moral decision procedure dominated by rules. But act-consequentialists deny rules have a role in the criterion of moral rightness. Prescribing a decision procedure dominated by rules and then denying rules a role in the criterion of rightness can be problematic. Rule-consequentialism gives rules roles first in the decision procedure agents should use and second in the criterion of moral rightness. But giving rules this second role has attracted objections, some of which are outlined and answered here. The final section of the paper considers some recent developments.

Key words

Act-consequentialism, Rule-consequentialism, Rules, Decision Procedure, Criterion of Rightness, Blame, Incoherence, Reflective Equilibrium, Collective Action Problems, Partial Acceptance Problems.

Consequentialist ethics is best thought of as a family of theories. The fundamental principles of the theories in this family share a focus on consequences, but some theories in the family evaluate acts by their consequences and other theories do not but instead apply the consequentialist test only to other things such as rules or motives (see Portmore, this volume). The most familiar members of the consequentialist family are act-consequentialism

and rule-consequentialism. The most familiar form of act-consequentialism, namely maximizing act-consequentialism, holds that an act is morally permissible depending on whether there is some alternative act that would produce better (expected) consequences. Rule-consequentialism holds that an act is morally permissible depending on whether the act is permitted by the rules with the best consequences.

The chapter explains why nearly all members of the consequentialist family make use of rules. Rule-consequentialism is often accused of giving rules too much importance. That accusation will be assessed here, as will some criticisms of rule-consequentialism made by non-consequentialists. I will also address some recent contributions to the development of rule-consequentialism.

Before I turn to rules, I will assemble the building blocks of consequentialism. The next section picks up the concept of the good, as used by consequentialists.

Section 1: The Good

A fairly simple consequentialist theory is utilitarianism, according to which the consequences that matter are additions to or subtractions from aggregate utility. Philosophers typically take 'utility' to refer to well-being (i.e., welfare, personal good). There are various views about what is the best account of well-being, such as hedonistic accounts, desire-fulfilment accounts, and objective list accounts (Parfit 1984, 493–502). Here is not the place to explore the contest between these views (Hooker 2015; Fletcher 2016; Crisp 2016; Woodard 2019, ch. 5).

All versions of utilitarianism take the aggregate good to be a matter of the well-being of all, added together *impartially*, such that a benefit or harm to one individual counts for exactly the same as does the same size benefit or harm to anyone else. One of the appeals of

utilitarian versions of consequentialism is precisely that they take this impartiality to be essential to the aggregation of utility and thus built into the foundation of morality.

Although we can imagine a consequentialist theory that focuses exclusively on something other than well-being, nearly all consequentialist theories have accorded well-being central importance, if not sole importance. The live question has been not whether well-being matters non-instrumentally but rather whether anything else matters as well non-instrumentally. One answer stretching back perhaps as far as the formula "the greatest good of the greatest number" is that equality matters as well as well-being. Someone who thinks that the consequences that matter non-instrumentally are not only gains or losses in terms of well-being but also equality of well-being could be called a distribution-sensitive consequentialist (Scheffler 1982).

Distribution-sensitive consequentialists might be persuaded that, in addition to aggregate well-being, what matters non-instrumentally is not that everyone have the same level of well-being. There could be situations in which the worse off cannot be raised to the well-being level of the better off and yet the better off could be lowered to the level of the worse off. Some consequentialists think that, when the better off are lowered to the level of the worse off, there is an increase in equality but a decrease in goodness.

These consequentialists might nevertheless be distribution-sensitive. They might hold, for example, that instead of ascribing non-instrumental value to equality of well-being, we should give higher priority to a benefit for someone whose level of well-being is low than to the same size benefit to someone whose well-being is greater. This 'prioritarian' view is not committed to 'levelling down' the better off but exerts pressure in favour of raising the worse off up to higher levels of well-being (Parfit 1997, 202–21; 2012, 399–400). A different form of distribution-sensitive consequentialism holds that what matters non-instrumentally is

neither equality nor benefiting the worse off, but instead getting everyone above some threshold of sufficiency (Crisp 2003). Such a view has been dubbed "sufficientarianism."

Utilitarianism clearly calculates aggregate utility impartially; but do prioritarianism and sufficientarianism? True, such views calculate the good in an agent-neutral manner. In our prioritarian calculation of the overall good, you and I are to give extra weight to benefits to the worse off compared to the same size benefits to the better off, whether or not you or I are among the worse off (see Portmore, this volume). Likewise, benefits to those below the threshold of sufficiency matter more than benefits to those above, whether or not you or I is below or above that level. But, as I explain elsewhere, impartiality is more than agent-neutrality, and so it is not true that a sufficient condition for a consequentialist theory to be foundationally impartial is that the theory be foundationally agent-neutral.[1]

Some recent forms of consequentialism definitely abandon the aspiration to have a theory that is fundamentally agent-neutral. These agent-relative forms of consequentialism accord more weight to the welfare of *individuals specially connected to the agent* than to the welfare of individuals without such a connection, and more weight to *acts of the agent's* than to acts of the same kinds performed by other agents. Forms of consequentialism that, at the foundational level, accord more weight either to the welfare of individuals specially connected to the agent than to the welfare of others without such a connection and/or to acts of the agent's than the same kinds of acts performed by others thereby build agent-relativity into the foundation of morality.[2]

Consequentialism can have more plausible practical implications if it incorporates into its 'value theory' more than just well-being. Consequentialists might even propose that certain kinds of act can have intrinsic *moral* value or disvalue, and that such value must be

[1] See my 2010, 35–9 and my 2013, 723–24.
[2] Most notably, Portmore 2011. See also Portmore and Hurley, this volume.

counted when consequences are assessed.[3] Incorporating into consequentialism the postulation of intrinsic value or disvalue for different kinds of act can help consequentialism have more intuitively plausible consequences. The same is true of building agent-relativity into the foundation of consequentialism.

However, each postulation a theory makes, even if the postulation seems intuitively correct, is one more thing the theory assumes rather than explains. If a theory, armed with however many intuitively plausible postulates, can explain the rest of the terrain better than any rival theory, then making those postulates seems necessary in order to come up with the most plausible theory. On the other hand, if there some rival theory makes fewer postulates and yet can explain the rest of the terrain as well as the theory that makes a greater number of postulates, then that rival theory has greater explanatory power (since it is explaining just as much but on the basis of fewer postulates). Thus there is what we might call an argument from parsimony against not only forms of consequentialism that postulate intrinsic value or disvalue for different kinds of act but also forms that build agent-relativity into the foundation of morality. The argument from parsimony is that there is some other moral theory that is more parsimonious because it does not start off with such postulates or with agent-relativity at the foundational level and yet is just as coherent with our intuitive convictions about right and wrong. If there is such a more parsimonious theory, one that has fewer postulates but is equally good at cohering with our moral verdicts, then that more parsimonious theory has considerable advantage.

What might philosophers who hold that kinds of act do have intrinsic value or disvalue and that there are agent-relative values and disvalues to be incorporated into maximizing act-consequentialism reply? They might say we antecedently believe that, for example, intentionally killing an innocent person against her will has more intrinsic moral

[3] Parfit 1984, 26; Feldman 1992, 182–185 and 1997, 164–169; Portmore, 2011; and Parfit 2017a, 395–406.

disvalue than intentionally letting an innocent person die against her will. Likewise, they might say that we antecedently believe that, for example, your intentionally killing an innocent person against her will has more disvalue for you than do two other people's intentionally killing two other innocent people against their wills. If there does appear to be greater intrinsic moral value or disvalue in certain kinds of act, and if there does appear to be agent-relative values to address, then shouldn't we take these appearances as a significant counterweight to the argument from parsimony?

There is disagreement about the correct answer to that question. My own view is that, even if there are such appearances, parsimony seems to me a decisive consideration when we are choosing between two theories that are roughly equally good at cohering with what seems to us to be true. Furthermore, I agree with Woodard (2019, p. 87) that many claims made about agent-relative value do not seem correct. There does not seem to be more disvalue in your doing something bad than in two other people each doing something equally bad.

This chapter does not have room to argue to a conclusion that there is a more parsimonious theory that can agree with our intuitive convictions about what actions and kinds of actions are right or wrong. The focus in this chapter is on the role(s) of rules in consequentialism. The discussion of the role(s) or rules would become unwieldy if we try to include forms of consequentialism that postulate intrinsic value or disvalue for different kinds of act and forms that build agent-relativity into the foundation of morality. Thus, henceforth the chapter will ignore those forms, and concentrate on consequentialist theories that do not postulate intrinsic value or disvalue for different kinds of act and that are completely agent-neutral at the foundational level.

Section 2: Actual vs Expected Value

Another question is whether consequentialism is formulated in terms of *actual* consequences or *probabilities* of consequences (see Cohen and Timmerman, Bykvist, and Jackson, this volume). There is an obvious rationale for caring more about what *actually* happened than about what was *reasonable to predict* to happen. In terms of the impact on well-being, benefits or harms that were possible but not actual matter less than benefits or harms that were actual. And yet at the point of deciding what to do, agents are almost never absolutely certain what all the consequences of the different actions they might do would be. And so telling them to do what will in fact produce the best consequences is not very helpful.

What is the right way for agents to deal with uncertainty about what the consequences would be of different possible actions? The most familiar answer—at least where the uncertainty is empirical uncertainty about what will happen (as opposed to uncertainty about which moral principles are correct)—concentrates on expected value. Expected value is calculated by multiplying the value or disvalue of each possible outcome times the probability that this outcome occurs and then summing the products. Here is a highly simplified example of an expected value calculation of one item:

Item being assessed	Values of possible outcomes of item being assessed	Probability of possible outcome	Expected value of possible outcome	Total expected value of item being assessed
An act	One possible outcome of positive value +5	Probability 40%	2	Total expected value of item:
	Another possible outcome of disvalue -3	Probability 60%	-1.8	0.2

An obvious question is: what determines the probabilities? A highly subjectivist answer is that probabilities are determined by what the agent believes about likelihoods. A less subjectivist alternative is that probabilities are determined by what people in the agent's milieu believe about likelihoods. And even less subjectivist view is that probabilities are determined by the evidence *available* to people at the time, even if that evidence has not shaped the beliefs held by people at the time.

Whether consequentialism should be framed in terms of actual consequences or expected value remains an unsettled issue. I will return to the distinction between *actual-consequence consequentialism* and *expected-value consequentialism* when I discuss possible conflicts between a prescribed moral decision procedure and an act-consequentialist criterion of moral rightness. However, the next section is about what moral decision procedure consequentialism prescribes. Choosing the act that actually will produce the best consequences cannot be the prescribed decision procedure, since we typically cannot know which of the acts we might choose actually will have the best consequences. Expected value is more important than actual value "when we are deciding how to act…." (Parfit 2017b, 228)

Section 3: The Role of Rules as a Moral Decision Procedure

A moral decision procedure is composed of a belief about how best to make moral decisions and a corresponding behavioural disposition to choose in the specified way. Since people cannot know the act that actually will produce the best consequences, should people make their moral decisions by calculating the expected values of the alternative acts available and then choosing the alternative with the highest expected value? Nearly invariably, this is not how people should make their moral decisions, for reasons that I will explain.

Different kinds of ignorance can come into play, and finding out relevant information can be costly, if even possible:

1. Very often, people who have a choice to make may *not know the full range of their available alternatives*. Sometimes such information can be obtained, but obtaining it typically takes time and attention and imagination.
2. Even when people have this information, they may *not know what the possible consequences are* of some available alternatives. Again, figuring out what the possible consequences are can take time and other resources.
3. Even when people know what the available alternatives are and what each available alternative's possible consequences are, *the value of some of these possible consequences may not be known*. Yet again, figuring out what the value of a possible consequence would be might be costly in terms of time, mental effort, etc.
4. Finally, even when people know what the available alternatives are, what their possible consequences are, and what the value of these possible consequences would be, *possible consequences' probabilities may not be known*.

One is not fully equipped to calculate all the expected values of all possible consequences of all available alternatives unless one knows what all the available alternatives are, what all their possible consequences are, what the values of all these possible consequences would be, and what the probabilities are of all the possible consequences. A decision procedure consisting of calculating all the expected values and then choosing the alternative with the highest expected value will be impossible to implement before all the inputs to the calculations are in hand. Sometimes, some inputs cannot be obtained.

Even where all needed information can be obtained, obtaining the information might be very costly. Sometimes, the costs of obtaining it is greater than the difference in expected values of various alternatives. Suppose I am deciding between buying inexpensive bicycle A and inexpensive bicycle B for the sake of riding to work more quickly than I can walk. I

don't know what the expected value of purchasing A is, and the same is true for purchasing B. But I'm pretty confident the difference in expected values is relatively small. If I spend many months trying to decide which to buy, and I lose forty-five minutes extra time each work day walking to work while I am deciding which to buy, the cost to me of making the decision on the basis of a fully researched expected-value calculation is greater than the difference between purchasing A and purchasing B. Whenever the calculation costs exceed the differences at stake—either by a lot or by a little—a decision procedure consisting of calculating all the expected values and then choosing the alternative with the highest expected value is not cost effective.

The problems I have cited so far with a decision procedure that consists in expected value calculations come from ignorance and the difficulties and costs of overcoming it. These problems would be tremendously important even if, once we had the information needed to do the calculations, we were perfectly accurate calculators. However, we are not perfectly accurate calculators, even at the best of times. And we often have to make practical decisions in a hurry, in which case calculation mistakes are especially likely.

In addition, personal biases might distort our assessment of the values of possible consequences or the probabilities of those consequences. For example, our bias towards ourselves might lead us to underestimate the amount of possible harm to others of decisions we might make that would benefit us. Likewise, our bias towards ourselves might lead us to underestimate the probability of harms to others.

Moreover, nearly everyone knows personal bias can distort people's calculations in these ways. Now suppose that this knowledge about the influence of bias were combined with a shared belief that the decision procedure used by others for making moral decisions was to try to calculate the expected values of the different alternatives available and then choose the alternative with the highest expected value. If we knew that others' calculations

would be distorted by their personal biases and that others would physically hurt us, steal from us, or break their promises to us whenever they convinced themselves that doing such an act has higher expected value than not doing it, we would have little confidence that others would not physically hurt us, steal from us, break their promises to us, or lie. If we felt others would lie to us, break their promises to us, steal from us or physically hurt us whenever they convinced themselves that such an act has the greatest expected value, we would be wise to devote a large share of our attention and other resources to protecting ourselves and our property. And our willingness to embark on diachronic cooperation with others and to trust what others assert would be minimal.

The problems resulting from ignorance, the costs of calculation, the influence of bias, and the need for assurance are each enough to condemn expected value calculation as a decision procedure for everyday use. The aggregate of these problems makes an overwhelming case against such a decision procedure.

Let us refer to the decision procedure commitment to which would produce the best consequences, or the highest expected value, as the *optimific* decision procedure. Nearly all consequentialists agree that a decision procedure consisting solely of trying to calculate the expected values of the alternative available acts and then choosing the alternative with the highest expected value would not be the optimific decision procedure. Virtually all consequentialists agree that the optimific decision procedure would consist of dispositions to comply with multiple common-sense rules, including prohibitions and requirements.[4] This decision procedure would be more feasible, cost effective, and reassuring to others than a

[4] A fairly recent articulation of this idea can be found in de Lazari-Radek and Singer 2014, 312–13. Classic discussions are J. S. Mill's references to secondary principles (Mill 1861, ch. 2); Sidgwick's development of the idea that even act-utilitarians should regularly think in terms of "common-sense morality" (Sidgwick 1907, Bk III, and Bk IV, chs. III and IV); and R. M. Hare's acknowledgement of the role of "intuitive-level" thinking (Hare 1981, chs. 3, 8, 9). See also Woodard 2019, 195–200.

decision procedure consisting solely of trying to calculate the expected values of the alternative available acts and then choosing the alternative with the highest expected value.

The prohibitions that would most obviously be included in the optimific decision procedure are ones on physically attacking people or their property, stealing, promise-breaking, and lying. But the optimific decision procedure would not consist solely of dispositions to comply with 'negative' rules such as the prohibitions just mentioned. Another rule that would be part of the optimific decision procedure would be a 'positive' rule about doing good for others in general.

However, this rule of general beneficence cannot be as strong as the rules against assault, theft, promise-breaking and lying, or we would be back with the increase in danger and distrust that I just mentioned. And the rule of general beneficence that is part of the optimific decision procedure cannot require one *always* to maximize expected value as long as one does not infringe the prohibitions on assault, theft, promise-breaking, and lying. If the rule of general beneficence went that far, we would be back with the above problems concerning ignorance and getting beyond it that I mentioned above.

Should the rule of general beneficence take the form of requiring one to have a standing disposition to benefit others in general when such benefits *are obvious* and providing them would not involve assault, theft, breaking promises, or dishonesty? The answer is no for two different reasons. One of these reasons is that the rule of general beneficence cannot be so categorical that it always trumps duties to those with whom one has special connections (on which, see Jeske, this volume). The second is that there are limits on the amount of self-sacrifice that the rule of general beneficence can require (see Sobel and Archer, this volume). I will explain both of these reasons below.

People often say that parents with an intense interest in general good and little interest in their own children tend to have unhappy children. If what is said here is true, what explains

why it is true? Children need love, especially from their parents. Are parents therefore morally required to love their children? Consider this counterargument:

Premise 1: What can be morally required of one is limited to things over which one has control.

Premise 2: Love is not something over which one has control.

Conclusion: What can be morally required of one does not extend to love.

Both premises of this argument can be challenged. However, I will not do that here. Instead, I will point out that, even if love includes affection and affection is not entirely in one's control, what is in one's control is taking a special responsibility for and interest in one's children. Arguably, even more than affection, what children need is for someone who knows them well to take special responsibility for and interest in them, and to do so on a sustained basis.

If that is what children need, then the optimific decision procedure will include a rule that someone does this for each child. Not always but normally the people most disposed to take special responsibility for and interest in a child are the child's parents. This disposition needs reinforcing by the addition to the decision procedure of an injunction to take special responsibility for and interest in one's own children.

Admittedly, such a rule will often lead people to do what is best for the children when the time, energy, or other resources involved could have instead been used to help other people more. Nevertheless, in general and on the whole, the world will be a less miserable place if every child has some people taking special responsibility for and interest in him or her.

I am not at all suggesting that the world would be a happier place if *unrestricted* nepotism pervaded everyday decision making. Nepotism always conflicts with equality of opportunity and typically conflicts with efficiency, and for these reasons must be restricted. What the restrictions on partiality towards one's children should be is too big a topic to comment on here. The important point here is that the optimific moral decision procedure is one that requires people to take special responsibility for and interest in their own children.

Although children most obviously need nurturing and protection, nearly all of the rest of us need affection, attention, and support too, sometimes. Admittedly, some people are loners suited to life away from human contact and attention. But they are rare exceptions to the generalization that people are social beings. Moreover, arguably, having deep friendships is not only instrumentally valuable to people but also a constitutive element of human well-being.

Central to friendship is mutual affection. A natural concomitant of one's affection for someone is giving that person some degree of priority when one is allowed to do so. Giving priority to friends not only flows from affection for them but also bolsters their affection in turn. A moral decision procedure that forbids prioritizing friends even when one is allocating one's own resources would thus endanger friendship. And friendship is one of the things crucial for human well-being. This is a compelling argument for having a moral decision procedure that *permits* one to prioritize friends when one is allocating one's resources.

We can go further and argue that the importance of friendship is so great that the optimific moral decision procedure would contain a rule *requiring* one to give some degree of priority to one's friends when one is allocating one's resources. One benefit of such a requirement is that the requirement will help sustain friendships. Another benefit comes from

the degree of assurance that internalization of the requirement will give people that their friends will regularly be trying to do good for them.[5]

A comparison between alternative possible rules should not be limited to a comparison of the consequences of people's following the rules. People's acceptance of rules can have consequences that are not the result of actions these people do. For example, many people would find it distressing to be prohibited from taking a special interest in their children, and such distress could predate whatever actions are produced in compliance with the rules. And maybe there are many people for whom the only way to get themselves not to take a special interest in their children would be for them to expunge or at least supress their affection for their children. For many people, the loss of affection for their children would imperil their relationships with their children and demote one of the chief sources of purpose and happiness in life. As I have suggested above, people's loss of affection for their children would also be very bad for their children!

So far, I have pointed to the costs and benefits of being known to accept various rules, the psychological costs and benefits to the agent of accepting various rules, and the difficulties and costs of making decisions by various rules. Another kind of cost I think should be counted are what I call teaching and internalization costs for rules. Teaching and internalization costs are the costs to those who teach the rules and those to whom the rules are taught.

It is possible to learn a rule in the sense of memorizing it while not really appreciating the rule's meaning or having one's motivations and dispositions be shaped by the rule. That is not the kind of learning a rule we are interested in here. By far the most important effects of a rule occur if one's motivations and dispositions are shaped by the rule. In order to

[5] Brandt 1989, n. 22; Sidgwick 1907, 434–35; Pettit 1997, 97–102; Mason 1998, 386–93; Powers 2000, 239–54.

emphasize that the kind of rule acquisition under discussion here involves not merely cognitive belief but also motivational and dispositional elements, I am using the term "internalization of rules" instead of the term "learning of rules."

Presumably, the costs in terms of time and attention and effort of either teaching or internalizing a *greater number* of rules rather than a smaller number of rules is greater. Likewise, the costs are greater of teaching or internalizing *more complicated* rules rather than less complicated ones. And the costs of teaching or internalizing rules that *conflict with natural inclination* are also higher than the costs of teaching or internalizing rules that require or merely permit actions favoured by natural inclination.

To be sure, teaching and learning have their benefits as well as their costs. These benefits must be taken into consideration when we are trying to assess rules or decision procedures. One possible benefit of teaching is the sense of gratification the teacher can get. The benefits of learning are even greater. For example, those learning self-control, trustworthiness, and sensitivity to others' feelings will probably have their own life prospects improved immensely. Moreover, an increase in the number, complexity, or demandingness of the rules that get internalized may very well bring benefits to others.

Having identified different costs and benefits of the teaching and internalization of rules, we can run a cost–benefit analysis of different possible sets of rules. Such a cost–benefit analysis will come out favouring a set of rules that includes prohibitions on physically hurting others, stealing, breaking promises, and dishonesty, plus a requirement that one take special responsibility for and interest in one's children and friends, plus a general requirement that one do good for others when one can do so without physically hurting others, stealing, breaking promises, or dishonesty. That is not all. There can be cases in which disaster threatens unless one infringes prohibitions on physically hurting others, stealing, breaking promises, and dishonesty, or unless one goes against the interests of one's children

and friends. So, in order to prevent disasters in such cases, the optimific decision procedure would include an overriding prevent-disaster rule (Hooker 2000, 133–36, 165–69).

But there will be limitations on the number, complexity, and demandingness of these rules. The reason for this is that the increased costs of teaching or learning ever more rules, or every more complicated rules, or ever more demanding rules will, at some point, outweigh the increase in the benefits that would be produced the internalization of these rules. Too *many* moral rules are too hard to learn and keep straight. The same is true of too *complicated* moral rules. In contrast, the problem with teaching and internalizing rules that are too *demanding* is not cognitive but motivational. Because people naturally care far more about themselves and special others than about strangers, getting people to internalize rules demanding self-sacrifice for the sake of strangers is difficult and gets more difficult as the level of likely demands climbs.[6]

How might such a limit on demandingness be formulated? I propose that both the rule about doing good for others and the rule about preventing disaster *can* require self-sacrifice that is significant but do not require sacrifice that is more than significant, though these rules do *permit* self-sacrifice beyond this level. Any very specific demarcation of the limit of self-sacrifice that can be required seems to me likely to be insufficiently flexible. So my own view is that, while "as long as the sacrifice over a whole life is at least significant" is vague, that vagueness is the price to be paid for needed flexibility.

Could someone be so badly off throughout her whole life that she is never required to make sacrifices for others, because it is never the case that the cost to this person of a sacrifice would be less than the benefit she could give to someone else? To allow for this possibility, I referred to self-sacrifice that *can* be required. Is there a point at which someone made enough sacrifice for others that he is justified then "to shut the gates of mercy on

[6] Cf. Wolf, 2016, fn. 18.

mankind"?[7] I cannot see how there can be a limit to the demands of self-sacrifice that morality can reasonable require without there being some point where refusing to make further sacrifices is justified.

As I wrote earlier, nearly all consequentialists accept that the optimific decision procedure for making every-day moral decisions is not to try to calculate the expected values of the different possible actions and then to choose the action with the highest expected value. The optimific decision procedure instead contains rules prohibiting certain kinds of action and rules requiring other kinds, with actions that are neither prohibited nor required then being optional. While all consequentialists I know of agree broadly about what these prohibitions and requirements are, I should not leave the impression that all consequentialists agree exactly what the correct test is for decision procedures.

Act-consequentialism is *individualist* about this matter in the sense that it prescribes *to you* the decision procedure whose internalization *by you* will produce the best consequences and *to me* the decision procedure whose internalization *by me* will produce the best consequences. If there is any difference between us or our circumstances, then perhaps the decision procedure whose internalization by you will produce the best consequences is not the same as the decision procedure whose internalization by me will produce the best consequences. There might thus be variation between us in how act-utilitarianism tells us to make our moral decisions.

Rule-consequentialism, in contrast, is collectivist in the sense that tests a moral decision procedure by the consequences of that decision procedure's internalization *by* everyone. Rule-consequentialism prescribes to the collective of us all the moral decision procedure whose internalization by the collective will have the best consequences, or highest expected value. And, because rule-consequentialism prescribes the same moral decision

[7] Here I borrow words from Thomas Grey's "Elegy Written in a Country Churchyard."

procedure for all, it gives to all of us the same guidelines and goals in our moral thinking. Indeed, rule-consequentialism seems to picture morality as a shared, collective enterprise. And the theory fits together smoothly with the idea that morality's requirements, permissions, and prohibitions should be suitable to serve as *public* rules—to be such that "general awareness of their universal acceptance would have desirable effects", to borrow words from John Rawls (1971, section 23).[8]

4. Rules as Part of the Criterion of Moral Rightness?

The term "criterion of moral rightness" is a useful term of art (Bales 1971). A criterion of moral rightness is a complete account of the properties of acts that make acts morally right. Likewise, the criterion of moral wrongness is a complete account of the properties of acts that make acts morally wrong. Rather than refer to criteria of rightness and wrongness, I will abbreviate as appropriate.

Act-consequentialism is the view that an act is morally right if, only if, and because of its consequences compared to the consequences of alternatives to this act. The most familiar form of act-consequentialism requires acts that *maximize* the impartial good, or that *maximize* expected impartial value. A less demanding version of act-consequentialism frames the theory in terms of *satisficing*, that is, in terms of bringing about enough good, even if less than the maximum.[9] I henceforth focus on maximizing versions of act-consequentialism.

What exactly are the maximizing act-consequentialist criteria of rightness? The term "morally right" is ambiguous. The term could mean "morally required" or it could mean "morally allowed" ("morally permitted"). This ambiguity has little importance *within* maximizing act-consequentialism. The theory holds that nearly all morally permitted acts are

[8] The publicity condition is implied in Baier's 1958, 195f and prominent in Gert's 1998. I've defended publicity before in my 2010, 111-17, and 2016, 145–49.
[9] See Slote 1984, 1985, and 1989. For compelling arguments against satisficing consequentialism, see Mulgan 2001, 129–42. Cf. Chappell, this volume.

also morally required. The only way, according to act-consequentialism, an act can be permitted but not required is if, in addition to this act, there is at least one other act available to the agent in the circumstances that would produce exactly as much good and no other act that would produce more good than either of these two acts would. Such circumstances presumably are fairly rare.

Many of the moral philosophers who are committed to act-consequentialism now call themselves global consequentialists. By this, they mean that they hold the more general thesis that any act, decision procedure, rule, motive, practice, institution, and so on is morally right if, only if, and because no alternative to that act, decision procedure, rule, and so on has greater (expected) value. Certainly, all global consequentialists are act-consequentialists. And, as I said, many act-consequentialists are global consequentialists. But I will explain how it is possible to be an act-consequentialist without being a global consequentialist.

When an act-consequentialist turns to the question of which decision procedure, rule, or motive is best, there is the question whether the best one is whichever will lead to the greatest number or percentage of good-maximizing acts or whichever will result in the greatest value (Frankena 1988). For the sake of illustration, consider two possible decision procedures, each of which leads to 1000 acts. The first of these decision procedures would lead to 990 good-maximizing acts but 10 disastrous acts. The second of these decision procedures would lead to 1000 acts each one of which is only a little sub-optimal. The second decision procedure might well produce greater value even though it produces zero good-maximizing acts.

An act-consequentialist who thinks that the best decision procedure is whichever one leads to the greatest number or percentage of good-maximizing acts is not a global consequentialist (for more on global consequentialism, see Greaves, this volume). An act-consequentialist who thinks that the morally right decision procedure is whichever one

produces the greatest value is a global consequentialist. Of these two versions of act-consequentialism, the version that focuses on the greatest value, not the highest number or percentage of good-maximizing acts, seems truest to the spirit of the view. No wonder most act-consequentialists take this path.

The distinction between the two versions of act-consequentialism I have just been discussing might seem a point about theory rather than practice. These two versions of act-consequentialism probably agree *in practice* on what the elements of the optimal decision procedure is. They will agree on this if they think the very same elements will both produce the greatest value and result in the higher number and percentage of good-maximizing acts. Nevertheless, the distinction is worth making in order to show how an act-consequentialist can avoid being a global consequentialist.

One thing that all versions of *act*-consequentialism agree about is that the role that rules appropriately play is in the recommended decision procedure, not in the criterion of moral rightness or in the criterion of moral wrongness. The act-consequentialist criteria of moral rightness and wrongness make no reference to rules: an act is morally right if, only if, and because no other act has higher value. Act-consequentialism holds that one can use the morally best decision procedure and yet be led by it to select a morally wrong act, as I will now illustrate.

Suppose you are asked by your boss whether you left work early yesterday. The decision procedure that consequentialism prescribes includes a rule against lying. Admittedly, it also includes a rule requiring you to prevent disaster, and that rule will sometimes conflict with and outweigh the rule against lying, but only where the only way to prevent disaster is to lie. Suppose that in the case at hand you cannot see that lying to your boss is needed to prevent a disaster. Thus, you tell the truth that you left early from work yesterday.

Now suppose that in fact your telling the truth to your boss in this case does not maximize the impartial good, because the truth that you reveal annoys your boss and creates tension between the two of you. So the decision procedure that consequentialism tells people to follow led you to do something that did not actually maximize the impartial good. I mentioned earlier that consequentialist theories can be formulated in terms of actual consequences or in terms of expected value. According to actual-consequence act-consequentialism, your act of truthfully answering your boss's question was not morally right, because it did not, as things turned out, produce as good consequences as your lying would have.

Was your act morally right according to expected-value act-consequentialism? We earlier discussed many impediments to the calculation of expected value. Definitely, in the few seconds between your boss's asking you whether you left work early yesterday and your having to answer or refuse to answer, you couldn't conduct an expected value calculation that takes into account every possible consequence of telling the truth, every possible consequence of lying, and every possible consequence of refusing to answer. But let us set all these difficulties aside and imagine that you could in fact quickly calculate the expected values of telling the truth, lying, and refusing to answer. And suppose you know your boss and your co-workers were elsewhere yesterday afternoon and so couldn't have seen that you left early. So you are very confident that you wouldn't be found out if you told a lie now. Suppose you are also nearly certain that your boss's hearing the truth will annoy her and create tension between the two of you. Thus, your expected value calculation comes out favouring lying to her.

Given that consequentialism advocates a moral decision procedure dominated by multiple rules such as I suggested above, then, where the expected values of acts that comply with those rules are lower than the expected value of acts that infringe those rules, the

prescribed moral decision procedure leads to an act that is morally wrong according to expected-value act-consequentialism. In such cases, the fact that rules play a role in the decision procedure that act-consequentialism prescribes but not in expected-value act-consequentialism's criterion of rightness yields practical conflict between the prescribed decision procedure and the theory's criterion of rightness.

Admittedly, conflict between prescribed decision procedure and criterion of rightness is likely to arise less often for expected-value act-consequentialism than for actual-consequence act-consequentialism. There are many cases where an act that violates the rules is committed but is not found out. In many of these cases, the consequences of the act would *actually* be better than the consequences of not doing the act. However, when an act that violates the rules is found out, the consequences are typically extremely negative, including blame, loss of trust, withdrawal of good will towards the perpetrator, and perhaps punishment. Even if the probability of the perpetrator's being caught is low, the negative consequences of being caught are typically so high that an expected value calculation of doing the act in question will often come out against doing the act. There are plenty of instances of rule violations that *turned out* to be good maximizing but an expected value calculation done *in advance* would have opposed.

Rule-consequentialism is the view that an act is morally wrong if, only if, and because it is forbidden by the code of rules whose widespread internalization would produce the greatest expected value. Rule-consequentialism broadly agrees with act-consequentialism about the kind of moral decision procedure people should have, one in which multiple rules predominate. However, unlike act-consequentialism, rule-consequentialism holds that multiple rules have an ineliminable role to play in the criteria of moral rightness and wrongness: acts are right or wrong depending on whether they are forbidden by *rules* that pass a consequentialist test.

Concerning many situations, rule-consequentialism and act-consequentialism will be in agreement about which available acts would be morally required, morally optional, or morally wrong. Whenever act-consequentialism holds that telling the truth or keeping a promise or leaving other people's property alone would maximize the good, rule-consequentialism agrees that such acts are morally required. Nevertheless, act-consequentialism and rule-consequentialism disagree about *why* such acts are morally required. Act-consequentialism maintains that they are morally required simply because these *acts* maximize the good. Rule-consequentialism holds that these acts are morally required because the optimific *rules* require them.

5. A Problem with Having the Decision Procedure and the Criterion of Rightness Conflict

Problems for act-consequentialism arise from the possible conflicts between its criterion of rightness and its decision procedure. On the occasions where agents follow the decision procedure that act-consequentialism prescribes but the act selected by this decision procedure is wrong according to the act-consequentialist criterion of moral wrongness, what judgement does act-consequentialism reach? Perhaps that seems like a misguided question. After all, the question itself states both that the agents follow the decision procedure act-consequentialism prescribes and that the act is wrong according to act-consequentialism. But such a statement seems to pose the question of whether act-consequentialism can, without an air of paradox, condemn an act that was selected by precisely the decision procedure that act-consequentialism tells agents to use. Surely, to maintain that an act is morally wrong is to condemn it. Yet, should an act be condemned if the agent who chose it followed the appropriate procedure in choosing it?

The act-consequentialist response to such queries is to insist that different questions get different answers. Which decision procedure should be used? The act-consequentialist answer is the decision procedure which produces the most good. What acts are morally right? The act-consequentialist answer is whichever ones actually maximize the good or at least have the highest expected value. Should agents who faithfully follow the prescribed decision procedure but choose an act that does not actually maximize the good or at least have the highest expected value be blamed? Act-consequentialists point out that this question about blame is different from the question of whether the agent followed the prescribed decision procedure and different from the question of whether the act thus selected maximized value. And the answer act-consequentialists usually give is that an act or agent should be blamed if and only if doing so will maximize the good. [10]

The problems with this answer are twofold. First, conceivably, the act arrived at by the prescribed decision procedure is one for which blaming the agent would somehow maximize the good. But does it make sense to blame someone for doing an act arrived at by the very decision procedure the agent is told to use? Second, conceivably, an act that maximised the good might also be one for which blaming the agent would maximise the good. Concerning such a case, the usual act-consequentialists' answer it that blaming the agent is recommended by their theory even if the act for which the agent is to be blamed was also morally right according to their theory.

The conjunction of the act-consequentialist thesis about how agents are supposed to make their moral decisions, the act-consequentialist thesis about which acts are right or wrong, and the act-consequentialist thesis about when blaming agents is appropriate is

[10] A different view might be that we should distinguish between having the attitude of blame and the act of expressing blame, and that attitudes are involuntary and so beyond act-consequentialist assessment. I can see how such ideas might be developed. But it seems to me that we do decide not only to express various reactive attitudes such as blame but also to have them. Be that as it may, the orthodox act-consequentialist position is that not only the expression of blame but also the having of this attitude can be assessed in act-consequentialist terms.

extremely counterintuitive. Suppose an agent follows the decision procedure that act-consequentialism tells the agent to use and, on this basis, decides to do X. Suppose X happens also to be the morally right act in the circumstances, according to the act-consequentialist criterion of rightness. Still, act-consequentialism maintains that whether it is appropriate to blame this agent for doing X is an open question. Admittedly, this conjunction of theses is not inconsistent within act-consequentialism. Nevertheless, it is extremely counterintuitive.

How can the problem I identified be avoided? Well, first of all, a tight conceptual connection between wrongness and blameworthiness must be maintained. True, the concepts wrongness and blameworthiness do not have exactly the same extension: an act can be wrong without being blameworthy, as in the case of an agent's act of throwing a punch after someone else has secretly drugged her without her knowing it. Nevertheless, unless the agent has an excuse for doing what was wrong, a wrong act is blameworthy. And an act cannot be blameworthy without being wrong.[11]

One way for act-consequentialists to escape this line of objection is for them to move from the view that whether it is appropriate to blame an agent depends upon whether blaming the agent has the best consequences to the view that whether it is appropriate to blame an agent depends upon whether the agent did something wrong without adequate excuse. Such a move would eliminate possible conflict between the act-consequentialist criterion of wrongness and the act-consequentialist account of blame. But the move would do nothing to address the possibility that an agent could faithfully employ the moral decision procedure that act-consequentialism prescribes and yet do what is wrong according to the act-consequentialist criterion of wrongness, unless faithfully following the correct decision procedure counts as an adequate excuse.

[11] Compare Mason, this volume, and Woodard 2019, chs. 3, 9.

Act-consequentialists would not be wise to try to escape from that difficulty by changing from prescribing a moral decision procedure consisting of the multiple rules we discussed above to prescribing a moral decision procedure consisting of expected value calculations. As we have already seen, there are compelling reasons why having a moral decision procedure consisting of expected value calculations is impractical and self-defeating.

Another way of responding to the problem of possible conflicts between the moral decision procedure and the criterion of wrongness is to amend the criterion of wrongness so that they more closely match the prescribed moral decision procedure. Since rules predominate the prescribed moral decision procedure, likewise they would have to predominate the criterion of wrongness. But to respond in this way is to abandon act-consequentialist criteria of rightness and wrongness and move to rule-consequentialism. Rule-consequentialism takes rules to predominate the criterion of rightness, as we saw at the end of section 4 above.

6 Is Taking Rules to be Part of the Criterion of Rightness a Mistake?

The foundation of rule-consequentialism is its principle about deontic status: whether an act is morally required, optional, or wrong depends on what acts are required, permitted, or forbidden by the rules with the best consequences. So the foundational principle of rule-consequentialism, its criteria of right and wrong, accords to consequences and rules ineliminable roles. This principle builds rules into the criteria of rightness and wrongness.

Consider a case where doing what is required by a rule selected for its consequences neither has as great expected value as some other act nor would actually have the best consequences. An example might be an act of keeping a promise. If rule-consequentialism nevertheless tells the agent to keep the promise, is rule-consequentialism being true to its nature as a consequentialist theory? If what all forms of consequentialism care most about is

making the consequences as good as possible, then any form of consequentialism that requires an agent to comply with some rule when this act would not produce the best consequences is incoherent.[12] This implies that the act-consequentialist criterion of rightness is the only criterion of rightness that coheres with consequentialism.

This objection might be telling if our route into rule-consequentialism started with an overriding moral commitment to bring about the best consequences. If rule-consequentialism sometimes reaches a conclusion about which act is morally required where this act is not the one that would maximize impartial good, then that conclusion would indeed fail to cohere with a starting premise that what is most important is maximizing impartial good. But we might have an argument for rule-consequentialism that has no consequentialist premise, much less an overriding moral commitment to bring about the best consequences.

For the avoidance of doubt, let me be crystal clear that I'm distinguishing between the foundational principle of rule-consequentialism and an argument for, or route into, rule-consequentialism. The foundational principle of rule-consequentialism does indeed assess acts by rules that are selected for their consequences. The argument for, or route into, rule-consequentialism, however, sits dialectically prior to that foundational principle. And this argument might have no consequentialist premise.

One prominent argument for rule-consequentialism is that it does better than any other moral theory at supplying a fundamental impartial principle that achieves a reflective equilibrium with our more specific moral intuitions about what kinds of act are right or wrong and about what particular acts are right or wrong.[13] This reflective equilibrium

[12] I first addressed this objection in my 1995. There I proposed that a rule-consequentialist agent's most basic moral motivation could be a concern for impartial defensibility rather than concern for the impartial good. See also my 2000, ch. 4; 2007, 514–19; and 2008, 75–85.

[13] For the methodology of seeking reflective equilibrium in ethics, see Rawls 1951 and 1971, 19–21, 46–51. The point that rule-consequentialism accords with our intuitions better than act-consequentialism was first made by Harrod 1936 and Urmson 1953. On the reflective equilibrium argument for rule-consequentialism, see my 2000, 4–30. For a subtle and sophisticated development of rule-consequentialism, see Mulgan 2006, ch. 5.

argument for rule-consequentialism has no consequentialist premise, which is one of the attractive features of this argument. Any argument for rule-consequentialism that employs a consequentialist premise would strike any non-consequentialist as question begging.

Having seen that there is an *argument* for rule-consequentialism that has no consequentialist premise, much less a premise that makes an overriding commitment to maximize the good, we now should consider whether a rule-consequentialist *agent* would have an overriding commitment to maximize the good. I indicated above that one rule a rule-consequentialist agent should accept is a requirement to do good for others in general if possible, at least up to some limit of aggregate self-sacrifice. But this rule does not have overriding force. If it did have overriding force, we would be back with most of the problems we saw with having maximizing the good as one's decision procedure.

Another rule that a rule-consequentialist agent should accept is one to *prevent disasters* up to some limit of aggregate self-sacrifice. The inclusion of a rule requiring the agent to prevent disasters is motivated on rule-consequentialist grounds (think of the benefits produced when people comply with this rule). This rule springs into operation when disasters threaten even when preventing the disaster is possible only if the agent breaks some other rule. So if breaking a promise is necessary in order to prevent a disaster, then rule-consequentialism tells the agent to break the promise.

A different objection from the one that rule-consequentialism is incoherent is the objection that rule-consequentialism reaches implausible verdicts about what to do. Suppose that the only way to prevent some disaster—such as a death or long-lasting misery—would be for the agent to break a promise of less than momentous proportions. In this sort of case, any moral theory that insisted that the promise be kept though keeping it will get in the way of preventing the disaster is counterintuitive. So if rule-consequentialism compels the agent to comply in such a case with the rule against breaking promises, then the theory is

counterintuitive. Such an objection to rule-consequentialism can come from act-consequentialists, whose own theory would not require an agent to keep a promise when breaking it would produce better consequences. The objection can also come from those non-consequentialists who are not absolutists about promise keeping.

However, this objection to rule-consequentialism is misguided. The inclusion of an overriding prevent-disaster rule in the set of rules prescribed by rule-consequentialism inoculates the theory against the objection that it would insist on compliance with rules even when disaster results.

The overriding commitment to *prevent disasters* must not be confused with an overriding commitment to *maximize the good*. There are cases where breaking a promise, telling a lie, stealing, or another kind of act condemned by one of the rules would produce slightly better consequences than not doing so would. In such cases, an overriding requirement to maximize the good would insist that the other rule be broken. An overriding commitment to prevent disasters would not have the same result, apart from in the extremely exceptional cases where breaking the other rule would prevent disaster and not breaking it would not, even though the difference in how good the consequences would be of the two acts would be only slight.

The objections that rule-consequentialism must be incoherent and that it might get in the way of preventing a disaster are therefore misguided. These two old objections to rule-consequentialism do not provide sufficient grounds for rejecting the rule-consequentialist position that assigns to rules a central role not only in the moral decision procedure agents should use but also in the criteria of rightness and wrongness. That is not to say that the rule-consequentialist position is not mistaken for some other reason, either old or new.

7 Recent Developments of Rule-consequentialism

Another old objection to rule-consequentialism is that it makes implausible demands on the agent when others are not following the optimific rules. Suppose one of the optimific rules requires all of us to restrict our behaviour in a certain way for the sake of the environment. Suppose doing so is burdensome to the individual, although of course everyone is better off if the environment is preserved than if it is spoiled. Now suppose you notice that others are in fact not restricting their behaviour. The objection is that rule-consequentialism requires you to follow the optimific rule even when this will be costly to you and do no good.

The objection that such a requirement is unfair was made by David Lyons (1965, 128–32, 137–42). The objection has been given a new twist and directed at Derek Parfit by Douglas Portmore (2017). Portmore contends that where others are not following optimific rules that are burdensome to the individual agent to follow, and where the individual agent's following these rules would not compel others to start following them and would not produce some other good, then there would be no sufficient reason for the agent to comply with the rule. If that is correct, then, in such a case, rule-consequentialism would be requiring the agent to do something that she has no sufficient reason to do.

This line of objection is very important, and perhaps especially as directed at Parfit's final position. Parfit's 2011 volumes argued that there are plausible forms contractualism that converge with rule-consequentialism in claiming that optimific rules determine right from wrong. Parfit's 2017a lays far more emphasis on two other arguments for rule-consequentialism. One is the reflective equilibrium argument (2017a, 421–22, 433, 450). The other is the argument about what explains the wrongness of acts that harm very many but harm each to only a very small or even imperceptible degree. This argument appeals "not to the separate effects of particular acts, but to the combined effects of what we and others together do. Some act would be wrong, we believe, if all optimific rules would condemn such

acts." (2017a, 432)[14] If Parfit ended up having as one of his two main arguments for rule-consequentialism that the theory handles cases where collective action is needed to prevent terrible aggregate harms, then his version of rule-consequentialism had better not make implausible demands in such cases. Parfit (2017b, 227, 229) replied to Portmore by repeating his contention that rule-consequentialism should be formulated such that the rules have to be optimific across different rates of acceptance by the population, and that one such rule would be "do not make sacrifices when these acts . . . would do no good whatever."[15]

Another recent development in rule-consequentialism can be found in recent papers by Susan Wolf and David Copp. Wolf outlines conceptualizing morality as a practice, "that is, as a loose and informal institution, itself perhaps embracing some smaller sub-institutions, constituted by a set of rules that specify 'offices, roles, moves, penalties, defences,' and so on." (Wolf 2016, 138) If one does think of morality in this way, then "one may without inconsistency or rule-worship admit that the point of the practice of morality is to bring about the greatest good for the greatest number without being committed to the idea that this is also the point of one's life." And Copp (forthcoming) conceptualizes the object of morality as being "to lead the society's members to be disposed such that they can live together with a minimum of conflict, meeting their needs and pursuing their values, cooperating with each other in joint projects that are important to them." If that is the point of morality, then Copp thinks an ideal moral code is one the currency of which would achieve this object at least as well as the currency of any alternative code.

The conceptualization of morality as a practice with the object Copp outlines is one that I accept. And I agree with Wolf and Copp that this conceptualization of morality goes hand in glove with rule-consequentialism. However, we cannot argue for rule-

[14] Compare Woodard 2019, ch. 5.
[15] On the problem of different rates of acceptance, see Tobia 2013.

consequentialism on the basis of this fit without first providing a compelling argument for conceptualizing morality in the way that Wolf and Copp do. The way to mount that argument seems to me to return us to the reflective equilibrium methodology. This conceptualization of morality fits better with our general and more specific beliefs about morality than rival conceptualizations do, or so I intend to argue in a forthcoming paper.[16]

[16] I am very grateful to Douglas Portmore for acute comments on an earlier draft of this paper. Many of the ideas canvassed in the revised version of the paper were ones he suggested that I discuss or at least admit.

Bibliography

Baier, K. (1958). *The Moral Point of View*. Ithaca, NY: Cornell University Press.

Bales, R. E. (1971). "Act-utilitarianism: Account of Right-making Characteristics or Decision-making Procedure?" *American Philosophical Quarterly* 8: 257–65.

Brandt, R, (1989). "Morality and Its Critics," *American Philosophical Quarterly* 26: 89–100.

Copp, D. (forthcoming). "The Rule Worship and Idealization Objections Revisited and Resisted," in M. Timmons (ed.), *Oxford Studies in Normative Ethics* 10.

Crisp, R. (2003). "Equality, Priority, and Compassion," *Ethics* 113: 745–63.

Crisp, R. (2016). "Well-Being," *Stanford Encyclopedia of Philosophy*.

Feldman, F. (1992). *Confrontations with the Reaper*. New York: Oxford University Press.

Feldman, F. (1997). *Utilitarianism, Hedonism, and Desert.* New York: Cambridge University Press.

Fletcher, G. (2016). *The Philosophy of Well-Being: An Introduction*. London: Routledge.

Frankena, W. (1988). "Hare on Levels of Moral Thinking," in D. Seanor and N. Fotion (eds), *Hare and Critics*. Oxford: Oxford University Press.

Gert, B. (1998). *Morality*. New York: Oxford University Press.

Hare, R. M. (1981). *Moral Thinking*. Oxford: Oxford University Press.

Harrod, R. F. (1936). "Utilitarianism Revised," *Mind* 45: 137–56

Hooker, B. (1995). "Rule-consequentialism, Incoherence, Fairness," *Proceedings of the Aristotelian Society* 95: 19–35.

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

Hooker, B. (2007). "Rule-consequentialism and Internal Consistency: A Reply to Card," *Utilitas* 19: 514–19

Hooker, B. (2008). "Rule-consequentialism versus Act-consequentialism," *Politeia* 24: 75–85.

Hooker, B., (2010a). "When Is Impartiality Morally Appropriate?" in B. Feltham and J. Cottingham (eds.), *Partiality and Impartiality: Morality, Special Relationships, and the Wider World.* Oxford: Oxford University Press, pp. 26–41.

Hooker, B. (2010b). "Publicity in Morality: Reply to Katarzyna de Lazari-Radek and Peter Singer," *Ratio* 23: 111–17.

Hooker, B. (2013). "Egoism, Partiality, Impartiality," in R. Crisp (ed.) *Oxford Handbook on the History of Ethics*. Oxford University Press, pp. 710–28.

Hooker, B. (2015). "The Elements of Well-being," *Journal of Practical Ethics* 3: 15–35

Hooker, B. (2016). "Wrongness, Evolutionary Debunking, Public Rules," *Etica & Politica* 18: 133–49.

de Lazari-Radek, K. and Singer, P. (2014). *The Point of View of the Universe*. Oxford: Oxford University Press.

Lyons, D. (1965). *Forms and Limits of Utilitarianism*. Oxford: Oxford University Press.

Mason, E. (1998). "Can an Indirect Consequentialist Be a Real Friend?", *Ethics* 108: 386–93.

Mill, J. S. (1861). *Utilitarianism* (many editions).

Mulgan, T. (2001). *The Demands of Consequentialism*. Oxford: Oxford University Press.

Mulgan, T. (2006). *Future People*. Oxford: Oxford University Press.

Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.

Parfit, D. (1997). "Equality and Priority," *Ratio* 10: 202–21

Parfit, D. (2011). *On What Matters*, Vol. 1, Vol. 2. Oxford: Oxford University Press.

Parfit, D. (2012). "Another Defence of the Priority View," *Utilitas* 24: 399–440.

Parfit, D. (2017a). "Responses," in S. Kirchin (ed.), *Reading Parfit*. Abingdon: Routledge, pp. 189–236.

Parfit, D. (2017b). *On What Matters*, Vol. 3. Oxford: Oxford University Press.

Pettit, P. (1997). "The Consequentialist Perspective," in M. Baron, P. Pettit, and M. Slote, *Three Methods of Ethics*. Malden, MA: Blackwell, pp. 92–174

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

Portmore, D. (2017). "Parfit on Reasons and Rule Consequentialism," in S. Kirchin (ed), *Reading Parfit*, Abingdon: Routledge, pp. 135–52.

Powers, M. (2000). "Rule Consequentialism and the Value of Friendship," in B. Hooker, E. Mason, and D. E. Miller (eds), *Morality, Rules, and Consequences*, Edinburgh University Press, pp. 239–54.

Rawls, J. (1951). "Outline for a Decision Procedure in Ethics," *Philosophical* Review 60: 177–97.

Rawls, J. (1971 and revised edition 1999). *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Scheffler, S. (1982). *The Rejection of Consequentialism*. Oxford: Oxford University Press.

Sidgwick, H. (1907). *Methods of Ethics*. London: Macmillan, 7th edition.

Slote, M. (1984). "Satisficing Consequentialism," *Proceedings of the Aristotelian Society Supplementary Volume* 58: 139–63

Slote, M. (1985). *Common-Sense Morality and Consequentialism*. London: Routledge and Kegan Paul.

Slote, M. (1989). *Beyond Optimizing*. Cambridge, MA: Harvard University Press.

Tobia, K. (2013). "Rule Consequentialism and the Problem of Partial Compliance," *Ethical Theory and Moral Practice* 16: 643–52.

Urmson, J. O. (1953). "On the Interpretation of the Moral Philosophy of J. S. Mill," *Philosophical Quarterly* 3: 33–39.

Wolf, S. (2016). “Two Concepts of Rule Utilitarianism,” in M. Timmons (ed.), *Oxford Studies in Normative Ethics* 6: 123–44.

Woodard, C. (2019). *Taking Utilitarianism Seriously*, Oxford: Oxford University Press.