Consequentialism and Reasons for Action[*]

Christopher Woodard

Abstract: Consequentialist theories often neglect reasons for action. They offer theories of the rightness or the goodness of actions, or of virtue, but they typically do not include theories of reasons. However, consequentialists can give plausible accounts of reasons. This chapter examines some different ways in which such accounts might be developed, focusing on Act Consequentialism and Rule Consequentialism and on the relationship between reasons and rightness. It notes that adding claims about reasons to consequentialist theories may introduce a welcome kind of complexity, and in doing so may help to make consequentialist approaches to ethics more appealing. For example, it may help consequentialists to explain the ideas of moral constraints and moral options.

Discussion of reasons for action is widespread in moral philosophy and in everyday ethical thought and talk. Reflecting on instances where someone has done something wrong, we might note that there was at least some reason for what they did. Wondering what to do, we may try to weigh up the reasons for and against some action. Yet consequentialist theories typically neglect reasons, focusing only on the rightness of actions. Or, if other topics are introduced, this is often a matter of extending consequentialist evaluation to other items, such as character traits or policies or institutions. With some notable exceptions, consequentialists often do not make explicit theoretical claims about reasons for action (exceptions include Crisp, 2006, ch. 2; McElwee, 2010; and Portmore, 2011). Since reasons for action are among the things we care about, and would like to understand better, it is worth considering what consequentialist approaches to ethics can say about them.

As I shall understand it, "consequentialism" is an approach to ethics which seeks to explain matters of ethical interest (such as the rightness of actions, or the justice of

[*] This is a draft of a chapter that has been accepted for publication by Oxford University Press in the forthcoming book *The Oxford Handbook of Consequentialism*, edited by Douglas Portmore.

institutions, or reasons for action) in terms of the goodness of outcomes. In discussions of consequentialism "outcome" is understood in an especially broad way, such that "the outcome of X" includes everything that would happen if X were realized, including the realization of X itself. So, for example, the outcome of an action includes the performance of the action (Portmore, 2011, 57). We should also understand "goodness" broadly, not building in more than is necessary to the definition of "consequentialism." Thus, let us say that consequentialism may employ an agent-neutral or an agent-relative account of goodness, and that it may understand goodness either as prior to reasons, or as a matter of what we have reason to desire (for discussion see Louise, 2004; Schroeder, 2007; and Portmore, 2011, ch. 3). Finally, note that consequentialism as I have defined it does not include a commitment to direct evaluation of anything. It allows for the possibility that what makes X favored (right, just, virtuous, or legitimate, say) is the value of Y (Kagan, 2000, 134–55).[1]

Let us begin by considering the concept of normative reasons for action, before exploring what Act Consequentialists might say about reasons, and then exploring what indirect forms of consequentialism might say about them. My aim will be to describe some of the theoretical options open to consequentialists who wish to give theories of reasons, and thereby to contribute to further exploration of those options.

1. Reasons for action

It is not possible to give an entirely uncontroversial characterization of the concept of normative reasons for action. I will not try to survey all of the controversies here.[2] Very broadly, I take the concept of a normative reason for action to be the concept of a consideration in favor of or against acting in some way.[3] This is to be contrasted, on one hand, with the concept of motivating reasons, which is the concept of a consideration that

[1] Many who discuss consequentialism take directness of evaluation to be a defining feature of it. This assumption appears to lie behind the worry that Rule Consequentialism is incoherent, for example.

[2] A recent collection of essays on reasons (including normative reasons for action) is Star, ed., 2018.

[3] In Scanlon's influential formulation, a normative reason in general is "a consideration that counts in favor of" something (Scanlon, 1998, 17; see also Crisp, 2006, 38; and Parfit, 2011, 31–3).

moves an agent to act by making it seem to her to be a good thing to do.[4] On the other hand, we can distinguish normative reasons for action from other normative reasons: there may be considerations in favor of or against believing something, desiring something, or hoping for something, for example.

One central controversy about normative reasons for action—"reasons for action" for short, from now on—is whether an agent's perspective in some way constrains what reasons she has. This controversy is connected to background views about the relationships between the concept of reasons for action and other concepts. For example, we might think that there is a close connection between the concept of reasons for action and the concepts of good deliberation, or of blameworthiness. If so, we are likely to think that what reasons for action an agent has is constrained in some way by her perspective, since good deliberation must connect in some way with the agent's perspective, and how things seemed to an agent is relevant to her blameworthiness. Alternatively, we might think that the concept of reasons for action is more closely connected to the concept of the rightness of actions than to good deliberation or blameworthiness. If so, the issue about perspective is left more open. On one view about rightness, the rightness of actions is not at all constrained by the agent's perspective. Something could then be a reason in favor of an agent acting in some way whether or not she could possibly be aware of it given her situation and her perspective.[5]

If we emphasize the connection with good deliberation or blameworthiness we are bound to think of reasons as in some way perspectival; if we emphasize the connection with rightness we can be more open-minded about whether they are perspectival, since we may or may not think that rightness itself is perspectival. In this chapter I will assume that there is a close connection between reasons and rightness, leaving aside the question of the connections between reasons, good deliberation, and blameworthiness.[6] This enables us also to set aside the issue of whether an agent's reasons are constrained by her perspective.

Consider, then, the relationship between the concept of reasons for action and the concept of rightness of actions. A reason is a consideration in favor of, or against, performing

---

[4] For a recent discussion of some different views about the nature of motivating reasons see Alvarez, 2018.

[5] For a defense of perspectivism see Lord, 2015. For criticism of one prominent argument for perspectivism see Way and Whiting, 2017. Some of these issues are explored in Woodard, 2019, ch. 3.

[6] Brian McElwee (2010, 397) claims that wrongness is more closely related to blameworthiness than to reasons.

some action. An action is right if and only if it is not wrong to perform it.[7] Intuitively, if there is a reason in favor of performing some action, that is certainly relevant to whether it is right to perform it. However, it does not seem to entail that it is right to perform it. For example, it may be wrong on some occasion to lie to a friend—even though doing so would protect his feelings, and this is a reason in favor of lying to him. Reasons for action may conflict with each other, so that there is a reason in favor of acting in some way but a stronger reason against acting in that way. Moreover, there may be more than one reason for or against any action. This suggests that the relationship we are after is between the rightness or wrongness of an action and the *overall set* of reasons for or against it.

Can we say more about this relationship? I shall assume, more specifically and controversially, that if there is *sufficient reason* to perform an action, then it is *not wrong* to perform it (which is to say that it is right, i.e. either required or optional). This leaves open a number of issues. First, it leaves open whether an action must be wrong when there is not sufficient reason to perform it. Perhaps it need not be: for example, perhaps there is no reason, and so not sufficient reason, to take *this* can of soup rather than its identical neighbor from the supermarket shelf, even though it is not wrong to take it.[8] Second, it leaves open what counts as "sufficient reason." One possibility is that there is sufficient reason to perform an action if and only if there is no weightier reason (or combination of reasons) in favor of any alternative. But, as we shall see later, that is not the only possibility. Third, it leaves open what explains the correlation between rightness and presence of sufficient reason. This could be because presence of sufficient reason makes it true that the action is right; alternatively, it might be that the balance of reasons, and the rightness of actions, are each to be explained in terms of the goodness of the outcome of the action.

Finally, the assumption that an action is right whenever there is sufficient reason to perform it leaves open the domain of reasons and the corresponding sense of rightness and wrongness. It is common to distinguish between kinds of normative reason for action. For example, we might distinguish moral reasons from legal or prudential reasons. We might also say that a certain action was legally right but morally wrong, for example. In the case of

---

[7] Standardly, "right" is taken to be ambiguous between "required" and "optional." An action is required if and only if it is wrong not to perform it. An action is optional if and only if it is not wrong to perform it, and not wrong not to perform it.

[8] There might be sufficient reason to take either can rather than none, of course. The example is drawn from Ullmann-Margalit & Morgenbesser, 1977, 761. See also Snedegar 2017.

moral rightness, an interesting further issue is whether the relevant domain of reasons is all reasons, or only moral reasons. Could there be most reason to do what is morally wrong?[9] In what follows I will leave these interesting issues about kinds of reasons aside, conducting the discussion instead in terms of a generic concept of normative reasons for action, and a sense of rightness that is left open deliberately. In this generic sense, an action is right whenever there is sufficient reason to perform it.

A reason for action, then, is a consideration in favor of or against performing some action. Which reasons an agent has may or may not be constrained in some way by her perspective. Her reasons may conflict with each other, and so she may have reasons to perform actions that it would be wrong to perform. But when she acts in ways for which she has sufficient reason, she does not act wrongly.

2. Act Consequentialism

Consider a simple form of Act Consequentialism, which says that an action is right if and only if there is no relevant alternative action with a better outcome. That is, this theory claims that there is a strict correlation between the rightness of actions and the goodness overall of their outcomes.

Reasons for action, we have assumed, may conflict with each other—so that there may be one or more reasons to do something that is wrong. There may also be more than one reason to do something that is right. The rightness of actions is related to the overall set of reasons, we assumed. According to Act Consequentialism, it is also related to the goodness overall of their consequences. If Act Consequentialism (and our assumption) is true, then, the rightness of actions is related both to reasons overall and to goodness overall.

This suggests a natural hypothesis: for Act Consequentialists, reasons are to be explained in terms of the different respects in which outcomes are good. Just as there may be a reason in favor of a wrong action, it is also the case that a wrong action may have an outcome that is good in some respect, even though it is not best overall. The contrast between goodness in some respect and goodness overall is a natural way for Act Consequentialists to relate the idea of there being a reason in favor of an action to the idea that it is right. According to this hypothesis, for Act Consequentialists rightness and wrongness correlate

[9] For discussion, see Scheffler, 1992, ch. 4; Portmore, 2011, ch. 2; and Forcehimes & Semrau, 2018.

with goodness overall, while individual reasons correlate with individual respects in which an outcome may be good. To return to our previous example: lying to your friend has an outcome which is good in one respect, since his feelings are not hurt; but if lying is nevertheless wrong, this outcome must be worse overall than the outcome of not lying to him.

On the definitions we have given, Act Consequentialism is compatible with a very wide variety of claims about which features of outcomes may be good in some respect.[10] This means that Act Consequentialists may claim for example that each token action of keeping a promise is in itself good. Thus they may claim that there is always a reason to keep promises. Or, to take a different example, they may claim that it is good that people get what they deserve, and thus that there is always a reason to give people what they deserve.

The hypothesis looks promising, but we must take care in spelling it out further. The difficulty is to spell out exactly what must be true of a feature of an outcome for its value to supply a reason to produce it, according to Act Consequentialism. One implausible answer to this question is that it is sufficient that the feature has positive value. This cannot be right. Suppose that everyone is currently blissful. You could either do nothing, in which case everyone's bliss would remain undisturbed, or you could press a switch, in which case everyone would be reduced to whichever level of well-being is minimally positive. It does not suffice for you to have a reason to press the switch that it would produce an outcome in which each person's well-being is positive. Intuitively, you have no reason at all to press the switch.

It is more promising to claim that a feature of an outcome supplies a reason to produce it if and only if it is better than *the corresponding feature* of some *relevant alternative outcome*.[11] In the case just examined, it is plausible to claim that the positive features of the outcome of pressing the switch—the barely positive level of well-being of each person—are all worse than the corresponding features of the relevant alternative outcome, where that means each person's level of well-being if you do not press the switch.

---

[10] Not just any theory of goodness can be combined with consequentialism. To avoid circularity, consequentialists cannot appeal to a theory of goodness that makes essential reference to the ethical phenomenon the consequentialist seeks to explain.

[11] This is to claim that whether a reason exists depends on the features of relevant alternative outcomes. This is compatible with, but does not entail, "contrastivism" about reasons. Contrastivism is the claim that the reason relation contains an argument place for alternatives. See Snedegar, 2017, 7–8.

If that is right, then this proposal correctly implies that you have no reason to press the switch, even though it results in an outcome with valuable features.

However, it is not trivial to provide a satisfactory general account of either of the elements of this proposal—that is, the idea of a "corresponding feature" and the idea of a "relevant alternative outcome." One difficulty attending the idea of a corresponding feature arises in "non-identity" cases. Suppose that you must do either A or B. If you do A, Tom will be born and will have a good life. If you do B, Tim will be born instead of Tom and will have an excellent life. Do you have a reason to do A? Tom's good life is a good feature of the outcome of doing A—but is it better than the corresponding feature of the outcome of doing B? That depends on whether we take the corresponding feature to be *Tom's* well-being in the world in which you do B, or instead to be *Tim's* well-being in that world. If a good life is better than no life, you have a reason to do A if we say that the corresponding feature is Tom's well-being, but not if we say that it is Tim's well-being (Parfit, 1987, ch. 16; Roberts and Wasserman, eds., 2009).

What about "relevant alternative outcome"? Act Consequentialists should probably say that the relevant outcomes are the outcomes of the agent's other options, where an option is an alternative action she could perform in the circumstances (see Smith, this volume; and Portmore, 2019). This seems to get many cases right. Suppose that Sylvie and Shreya are both badly off, and that you could either make Sylvie somewhat happy, or Shreya very happy. The best outcome overall is produced by making Shreya very happy, let us say, so according to Act Consequentialism that is the right thing to do. Nevertheless, intuitively there is some reason to make Sylvie happy. Her happiness is a consideration in favor of doing that. If you were to make her happy you would have acted wrongly according to Act Consequentialism, but there would have been a reason for acting in that way (for related discussion, see Chappell, 2015). Now, if we specify the "relevant alternative outcome" as the outcome of your other option, we can explain this reason. Sylvie's happiness is good compared with her state if you benefit Shreya instead. So, on the current proposal Sylvie's happiness supplies a reason to benefit her.

However, specifying relevant alternative outcomes in this way may seem to yield the wrong implications in other cases. Sometimes it is tempting to say that the relevant comparison is with the state of the world prior to action. Suppose that Lulu and Lisa are currently miserable. Suppose further that you could either make just Lulu happy, or both Lulu and Lisa happy, and that these are your only two options. Arguably, you have some reason to make just Lulu happy, even though it would be wrong to do so according to Act

Consequentialism. Her happiness seems to be a consideration in favor of making her happy. But we cannot explain this if we compare her happiness with her state if you were to take your other option, since she is equally happy in that outcome. Thus, we might be tempted to take Lulu's miserable state before you acted to be the relevant comparison.

Though it might seem intuitively correct in this case, Act Consequentialists should probably resist the temptation to specify the relevant alternative outcome in this way. One reason is that in other cases the comparison with the state of the world before the agent acts seems to give clearly the wrong answer (Norcross, 1997, 8–9). Another, more theoretical, reason is that it would mean that the Act Consequentialist account of rightness of actions (which *is* standardly defined in terms of the agent's options) would fall out of step with its account of reasons, since they would be based on comparisons with different sets of alternatives.

So we have the outline of an Act Consequentialist theory of reasons, but it leaves open some issues. The outline account is this: an agent S has a reason to perform some action A in circumstances C if and only if and because S could do A in C and the outcome of A in C would have some feature F which is better than the corresponding feature of a relevant alternative outcome.[12] As we have just noted, there are some difficulties in specifying what is to count as a corresponding feature, and which are the relevant alternative outcomes. In addition, of course, an Act Consequentialist must also give some account of goodness if she is to reach determinate conclusions about which reasons there are.

If an Act Consequentialist can answer these questions she can produce a theory of reasons for action. The theory would tell us which reasons an agent has in any specified circumstances, and why. Since Act Consequentialism already contains a theory of rightness of actions, if we also assume that an action is right whenever there is sufficient reason to perform it, she will also have specified the notion of sufficient reason. There is sufficient reason to perform an action, according to Act Consequentialism, if and only if there is no relevant alternative action with a better outcome. The most natural interpretation of this claim is as reflecting two underlying thoughts. One is that the strength of a reason is proportional to the value of the feature that supplies it, and the other is that there is sufficient reason to

[12] Should this read "better than the corresponding features of *all* relevant alternative outcomes"? No: if you could leave Tom in misery, or make him happy, or make him very happy, you have some reason to make him merely happy—even though it would be wrong to do so, according to Act Consequentialism.

perform an action only if it is favored by the strongest reasons overall. If we make these further claims, we get the correspondence between best overall outcomes and rightness of actions that Act Consequentialism asserts.[13]

3. Indirect Consequentialism

Now consider a standard form of Rule Consequentialism, according to which an action is right if and only if it is permitted by the best set of rules. Whereas Act Consequentialism asserts a correlation between the rightness of actions and the goodness overall of their outcomes, Rule Consequentialism asserts a correlation between the rightness of actions and the goodness of sets of rules. The specific correlation it asserts is that right actions are in every case permitted by the best set of rules, and that wrong actions are, in every case, prohibited by the best set of rules.

There are important issues arising from the question of how to evaluate sets of rules. Since rules do not, all by themselves, have causal consequences, they must be "embedded" in some way, which means drawing some appropriate connection between the rules and things with causal consequences (Kagan, 2000). In brief, there are two main issues. One is whether we should characterize the consequences of a set of rules in terms of the consequences of compliance with those rules, or instead in terms of the consequences of acceptance of those rules. Neither compliance nor acceptance entails the other, so we would expect different results depending on which way we go on this issue (Hooker, 2000, 75–80). The other issue is whom we take to comply with, or accept, the rules in order to characterize the consequences of the rules. There are many possible answers to that question, and many have been explored in recent discussion (for example, see Ridge, 2006; and Smith, 2010). At one extreme, we could characterize the consequences of a set of rules in terms of the consequences of the agent alone complying with (or accepting) them on a single occasion. At the other extreme, we could characterize consequences in terms of everyone everywhere always complying with (or accepting) them. The first extreme option would result in a theory

[13] Alternatively, Act Consequentialists could offer some account of "sufficient reason" that does not refer to goodness overall. If that were possible, then they could formulate their theory without mentioning goodness overall. Individual good features of outcomes would explain reasons, and rightness could be explained directly in terms of the notion of sufficient reason. This would provide one way for Act Consequentialists to respond to worries about the concept of goodness overall, such as those expressed in Foot, 1985.

much like Act Consequentialism, while the second extreme option would result in something more similar to Kantianism. Many possible versions of Rule Consequentialism lie between these extremes (Woodard, 2013).

In the case of Act Consequentialism, we explored the hypothesis that the fact that an outcome may have different good features explains how there can be multiple reasons, including reasons to perform wrong actions. Since Rule Consequentialism also appeals to the value of outcomes—albeit the outcomes of sets of rules—we could in principle explore the same hypothesis in relation to it. But a more obvious strategy is to try to explain the plurality of reasons for or against a single action in terms of the plurality of rules governing a single action. To return to our earlier example, we might say that there is a rule about lying, and a rule about looking after friends, which together explain why you have a reason to lie to your friend to protect his feelings, even though (we assumed) it would be wrong to do so. The rule governing lying is associated with a general reason not to lie, while the rule governing friendship is associated with a general reason to promote one's friends' well-being. (No doubt the best rules are more complex than this.) In this instance, these reasons conflict with each other, and it happens that the reason not to lie is stronger. It is natural to think of the rules that feature in the best set as associated with different considerations or reasons.[14]

Exactly how we should think of these rules depends on the way we answer the questions about embedding. If we formulate Rule Consequentialism in terms of compliance, the rules specify behavior. The consequence of complying with a rule that says "do not lie" is the consequence of people not lying. Which people? That is specified by our answer to the second embedding question. At one extreme, it is the agent alone on a single occasion. At the other, it is everyone at all times in all places. Rule Consequentialists may, but need not, believe that the embedding questions must be answered realistically. They might take a deliberately idealizing approach, because they think that it is part of the concept of morality that it consists of rules that form an ideal code, in the sense that things would go well if (more or less) everyone lived up to them (Hooker, 2000, 1 and 80–5).

Of course, we have not got a determinate outcome if we say only that it consists of people not lying. Not lying in which way? There are many different ways of not lying on each occasion on which to lie is an option, not to mention the many different ways of behaving when lying is not an option. Rule Consequentialists must say more if they are to characterize the consequences of rules. One thing they say is that the consequences to

[14] Brad Hooker (2000, 88–92) appears to think of them in this way.

consider are the consequences of the *whole set* of rules. So, if we are characterizing these in terms of compliance, the relevant outcome would be one in which the relevant agents comply with all of the rules at once. That is likely to be much more determinate. Indeed, we might then worry about the opposite problem, of it being impossible to comply with all of the rules at once. Rule Consequentialists must either tailor the rules so that they do not conflict—or, more likely, address the issue of what the set of rules requires when it is not possible to comply with all members of the set at once. This is likely to involve the idea that some rules take precedence over others.

However all of these questions are settled, compliance with a set of rules is ultimately a pattern of behavior. To comply is to behave in a way that the rule permits, and so compliance overall is a concatenation of pieces of behavior, by one or more agents. Suppose, as we have been doing, that complying with the best set of rules involves your not lying to your friend, even at the cost of hurting his feelings. What account can the Rule Consequentialist give of your reason not to lie on this occasion?

They might say, simply, that the source of the reason is the rule. But that invites worries about rule-fetishism. If they wish to go further, one thing they could say is that not lying on this occasion is *your part now* in the pattern of behavior that consists of all relevant agents complying with the best set of rules. This answer relies on a common idea, which I have elsewhere labelled the idea of "pattern-based reasons" (Woodard, 2013; and Woodard, 2019, ch. 5). The idea is that the fact that some action is part of a favored pattern of action can provide a reason to perform the action. In the current context, what makes the pattern "favored" is that it has better consequences than any alternative pattern performable by the same set of agents (namely, those specified in answer to the second embedding question). Assuming that this includes other agents, we could say that your not lying on this occasion is your part, right now, in the best that the whole set of agents could do. This is to treat the *parthood* relation between a token action and a favored pattern of action as providing a reason to perform the part. Schematically, the idea is this: agent S has a pattern-based reason to perform action A in circumstances C if and only if and because S could do A in C and A is S's part, in C, of some favored and eligible pattern of action P.

The idea of pattern-based reasons appears to be part of common ethical thought. It is common for people to explain their actions in terms of larger patterns of action of which they are parts, and which they take to be good or right (for some evidence, see Bardsley, Mehta, Starmer, & Sugden, 2010). Moreover, it is common for people to use the language of parthood to explain their thinking: people often say "I want no part in that," or "you should

play your part," for example. So, if Rule Consequentialists appeal to the idea of pattern-based reasons in their account of reasons for action, they are, at least, appealing to an idea with common currency. Ordinary ethical thought appears to recognize playing one's part as the source of a kind of reason, as well as recognizing the kind of reason that Act Consequentialism articulates, of causing good outcomes.

The idea of pattern-based reasons raises a number of theoretical puzzles, however. A central puzzle concerns the concept of "eligibility" of patterns. This marks the difference between those possible patterns of action which do, and those which do not, generate reasons to perform their parts. Presumably not every good possible pattern of action generates pattern-based reasons. What, then, distinguishes those which do (the eligible) from those which do not (the ineligible)? If we interpret Rule Consequentialism as employing the idea of pattern-based reasons, answers to the second embedding question (about which agents to specify, when characterizing the consequences of a set of rules) are, in effect, answering the question about eligibility. If we say that the consequences of the set of rules are the consequences of the agent on this occasion complying with those rules, we are in effect saying that the only eligible pattern is one consisting of this agent's action on this occasion.[15] If, on the other hand, we say that the consequences of the set of rules are the consequences of every agent on every occasion complying with those rules, we are in effect saying that this highly extended, highly idealizing, pattern of action is eligible. One question for those interested in this sort of theory is thus whether we can give satisfying explanations of why some patterns should be treated as eligible and others not.

A related question is whether pattern-based reasons, if they exist, would have any practical significance. This is related to the "collapse worry" about Rule Consequentialism (Hooker, 2000, ch. 4). In the current context, the issue is whether pattern-based reasons would make any difference to the total set of reasons agents have, or to which actions are right. The answer depends on how we specify eligibility. It is pretty clear that my part in the best that every agent could do might be different from the best thing I can do taking as given others' non-ideal behavior. On the other hand, an idealizing answer to the question about eligibility might be thought to purchase practical significance at the cost of plausibility.

[15] As this suggests, we can think of the reasons posited by Act Consequentialism as a limiting case of pattern-based reasons, in which the "favored pattern" is identical to the agent's act. See Bacharach, 1999, 118. We can also think of the actualism *vs*. possibilism debate in deontic logic in terms of pattern-based reasons: see Woodard, 2009.

Critics of idealizing theories ask why we should think that the fact that some pattern would have good consequences can provide a reason to play one's part in it, if this pattern would not be realized because others would not play their parts (Parfit, 2011, 312–20; Dietz, 2016; Podgorski, 2018). So an issue for Rule Consequentialists, on this way of understanding their view, is whether it is possible to specify eligibility in a way that is both plausible on general grounds and yields practical significance.

A further puzzle is how to recover the idea of plural, possibly conflicting, reasons if we appeal to the consequences of a whole set of rules. The consequences of compliance with a set of rules are the consequences of a pattern of behavior. But how do we get plural reasons from a single pattern? When you do not lie to your friend, you are (we have been assuming) playing your part in the pattern of behavior that constitutes compliance (by the relevant agents) with the best set of rules. So, we can use the idea of pattern-based reasons to offer an explanation of your reason not to lie. But what can we say about the other reason we supposed you have, to lie so as to protect your friend's feelings?

One option would be to treat this as a reason of the sort that Act Consequentialists recognize. Lying on this occasion would protect your friend's feelings, and that is a good feature of this outcome. This would be to think of Rule Consequentialism as a pluralist theory, recognizing both pattern-based reasons and the kind of reasons that Act Consequentialists recognize. This is, on independent grounds, a plausible interpretation of Rule Consequentialism, at least when it is specified in terms of compliance. For Rule Consequentialists tend to claim that the best set of rules includes rules instructing agents to do the best they can, as individuals, in the circumstances. One example of this is the "disaster prevention rule" that Rule Consequentialists tend to recognize (Brandt, 1992, 151; Hooker, 2000, 98–9).[16] This instructs agents to break any other rule when necessary to prevent a disaster. But this is not plausibly interpreted as a pattern-based reason: it is the good outcome of breaking the rule on this occasion that generates the reason, not the good outcome of a larger pattern of rule-breaking.

So Rule Consequentialists might explain plurality simply by postulating pattern-based reasons associated with a single pattern (which reasons could not conflict among themselves) together with the reasons that Act Consequentialists recognize (with which the pattern-based

[16] Another example is the rule requiring beneficence that Hooker postulates (2000, 98 n. 7). This is a general reason to benefit others, though it is subordinate to other rules in the ideal code.

reasons could conflict). But another possibility would be to evaluate the contribution of individual rules to the consequences of the best set. The obvious way to do this would be to compare the consequences of the best set with a series of alternative sets, each of which differs from the best set by removing a single rule at a time. Thus, for example, we get a sense of the contribution of the best rule governing lying by considering the difference between the consequences of compliance with the best set B, and the consequences of compliance with a different set of rules B* which lacks only that rule. We can then think of the rule governing lying as associated with a general reason against lying, which is a pattern-based reason to play one's part in the pattern specified by B rather than the pattern specified by B*.[17]

Finally, we should return to the first question about embedding—which was whether to characterize the consequences of a set of rules in terms of compliance with them or acceptance of them. The discussion so far has assumed that we go with compliance. This enables us to think of the consequences of a set of rules as being the consequences of a pattern of behavior, and thereby to employ the idea of pattern-based reasons. If instead we go with acceptance, we might have to give a different account of Rule Consequentialist reasons. As Hooker understands acceptance, for example, "to accept a code of rules is . . . to have *a moral conscience of a certain shape*. In other words, when rule-consequentialists consider alternative codes of rules, they are considering alternative possible contours for people's consciences" (Hooker, 2000, 91, emphasis in original). As this suggests, Rule Consequentialism of this sort is similar to Motive Consequentialism, which is another indirect form of consequentialism. According to a standard formulation of Motive Consequentialism, an action is right if and only if it would be performed in the circumstances by an agent with the best motives (Parfit, 2011, 375).[18]

Can we furnish Hooker-style Rule Consequentialism, or Motive Consequentialism, with a theory of reasons along the same lines as the one we considered for compliance-style Rule Consequentialism? That depends on whether the consequences of a moral conscience, or a set of motives, can be cashed out entirely in terms of a pattern of behavior. Such a

[17] This is to claim that whether a pattern-based reason exists depends on the features of the outcomes of relevant alternative patterns. This is parallel to Act Consequentialism's claim that whether a reason exists depends on the features of the outcomes of relevant alternative actions.

[18] Motive Consequentialism also faces issues of embedding. For relevant discussion, see Kagan, 2000; and Bradley, 2018.

pattern would not consist of compliance with the best set of rules, but of behavior generated by the best set of motives or the best conscience. This behavior is likely to include the behavior of people other than the agent with the motive or character. For example, if someone is highly disposed to break moral rules, this disposition might cause others to behave warily around him. If the consequences of the best conscience or set of motives can be cashed out in terms of behavior in this way, then we could apply the concept of pattern-based reasons in roughly the same way as we did with compliance-style Rule Consequentialism (albeit with reference to different patterns of behavior, which bear a different relation to the rules constituting the best set). If, on the other hand, the consequences of motives and consciences extend beyond the consequences of behavior, these forms of indirect consequentialism would have to find some alternative way of explaining agents' reasons for action.

4. Consequentialism and constraints

One important issue for consequentialist theories of reasons is whether they can account for all of the reasons that we believe exist. An interesting instance of that question is whether consequentialist theories of reasons can explain the existence of moral constraints.

The idea of moral constraints is that it is sometimes morally wrong to act in a way that has the best outcome overall, impersonally evaluated. This idea seems to be one component of the idea of moral rights. If Smith has a moral right to bodily integrity, it may be wrong for a beneficent surgeon to seize his organs to save five others, even when seizing them would make the outcome best overall, impersonally evaluated. Smith's moral right acts as a constraint in the sense that it constrains what may be done permissibly to promote the impersonal good (Scheffler, 1982; Kagan, 1989, 4, 24–32).[19]

One natural way to interpret the idea of moral constraints is to think of them as applying to *kinds* of action. For example, we might think of the kind of action *seizing someone's organs* as governed by a constraint, which applies in the case just mentioned. We can then distinguish between absolute and non-absolute moral constraints. An absolute moral constraint implies that every token action of the proscribed kind is morally wrong, while a non-absolute moral constraint fails to imply this. A non-absolute moral constraint might imply instead that there is a reason—perhaps a strong reason—not to perform actions of that

[19] Scheffler calls constraints "agent-centred restrictions."

kind (Kagan, 1989, 4–5). Note also that constraints need not be, in an intuitive sense, "negative" in the way that the constraint against seizing organs is negative. There could be a constraint requiring gratitude in response to kindness, for example. Such a constraint would imply that it is sometimes wrong not to express gratitude, even when necessary to make the outcome best overall, impersonally evaluated. Intuitively, we might describe this by saying that we are constrained to express gratitude (though we could equally describe this negatively, of course, by saying that there is a constraint against failing to express gratitude).

The idea of constraints is very important in many ethical views. As already noted, it is one component of the idea of moral rights. It is also one component of standard views of the moral force (in some circumstances) of legal rights. If you have a legal ownership of your driveway, it is at least sometimes wrong for others to park on it without your permission, even if doing so has the best outcome.[20] More broadly, many ethical views attribute significance to kinds of action in a way that the idea of constraints seems to capture. Ross's objection to Act Utilitarianism, for example, was that it (wrongly) implies that it is right to break a promise whenever doing so makes the outcome even a whisker better. A natural way of interpreting this objection is to think of Ross as claiming that there is a reason to keep promises (a reason against breaking them) because of the kind of action that *keeping promises* is—a reason which the utilitarian fails to take into account (Ross, 2002/1930, 34–5).

Of course, one option for consequentialists is to deny the existence of moral constraints. But it is worth considering whether their theories of reasons could accommodate them. Consider first whether Act Consequentialism can do so. Recall that, according to our outline account, Act Consequentialism claims that an agent S has a reason to perform some action A in circumstances C if and only if and because S could do A in C and the outcome of A in C would have some feature F which is better than the corresponding feature of a relevant alternative outcome. Let us suppose that action A is of the kind *keeping promises*. Act Consequentialism is compatible with very many theories of the goodness of outcomes, including those that claim that it is noninstrumentally good to keep promises. So, when combined with such a theory of goodness, Act Consequentialism will imply that there is always a reason to keep a promise, in virtue of the fact that keeping it will have an outcome with a feature (the fact that the promise is kept) which is better than the corresponding feature (the fact that the promise was broken) of a relevant alternative outcome (the outcome of breaking the promise).

[20] This example is drawn from Lyons, 1980, 17–28.

This goes some way to accounting for a constraint against breaking promises, but it is not yet enough. Act Consequentialism claims that an action is right if and only if there is no relevant alternative action with a better outcome. The idea of a constraint, we said, was that it is sometimes wrong to act in ways that make the outcome best, impersonally evaluated. Therefore, for Act Consequentialism to accommodate a constraint against breaking promises there must be some occasion on which a single token act of promise-breaking would have the following seemingly incompatible features:

i. Its outcome would be worse overall than the outcome of keeping the promise (this act would be wrong according to Act Consequentialism).
ii. Its outcome would be better overall, impersonally evaluated, than the outcome of keeping the promise (this act would violate a constraint against promise-breaking).

As this suggests, in order to explain constraints Act Consequentialists must distinguish between different senses of "better overall". Breaking the promise on this occasion must be better overall in an impersonal sense, according to the idea of a moral constraint. But it must also be worse overall, to be wrong according to Act Consequentialism. So, the sense in which it is worse overall must not be the (same) impersonal one.

The way that Act Consequentialists try to reconcile these claims is by introducing a non-impersonal, or "agent-relative," way of evaluating outcomes. They can say that what matters (at least some of the time) for an agent's reasons, and the rightness of her actions, is the agent-relative value of the outcomes of her actions, not their impersonal value. For example, they can say that, when it comes to keeping promises, actions of the kind *the agent herself breaking promises* are worse than actions of the kind *others breaking promises*. They can then say that it is worse overall, in this agent-relative sense, for her to break the promise, even though it is better overall, impersonally, for her to do so. By adopting a suitable agent-relative theory of goodness, Act Consequentialists can try to explain moral constraints (Portmore, 2011, ch. 4).[21]

Indirect consequentialists, such as Rule Consequentialists, can offer a different sort of explanation. Since their theories are not extensionally equivalent to Act Consequentialism, they anyway imply that it can be wrong to perform some action even when it would have the best outcome overall. That basic feature of the idea of moral constraints fits easily into the

[21] As is well-known, the theory of goodness would also have to be time-relative.

structure of indirect consequentialist theories. A more pressing question is whether it is plausible that the best set of rules, or the best conscience or set of motives, corresponds with the moral constraints that we wish to explain.

We can approach this question using the hypothesis that indirect theories employ the idea of pattern-based reasons. Recall that the idea is that an agent S has a pattern-based reason to perform action A in circumstances C if and only if and because S could do A in C and A is S's part, in C, of some favored and eligible pattern of action P. If we were to ask, for example, whether an indirect theory can explain a moral constraint against breaking promises, we could treat this as being the question whether it can explain the eligibility and value of some pattern P in which the agent's part in the relevant circumstances is not to break a promise. This pattern might be one in which no agent ever breaks a promise, or it could be something more complex. If a plausible explanation of the eligibility of some suitable pattern can be given, then an indirect theory could explain the idea that there is a reason not to break promises, even when doing so makes the outcome better.

The project of explaining constraints illustrates some basic theoretical choices facing consequentialist theories of reasons. Act Consequentialism has a simple structure, and so any explanation it gives of the complexity of reasons requires a complex theory of the value of outcomes. Rule Consequentialism, along with other indirect theories, has a more complex structure, and so it can retain a relatively simple theory of the value of outcomes while seeking to explain the complexity of reasons. Each approach has its own attractions and faces its own challenges.

For example, it is somewhat plausible to say that there is special disvalue in the agent breaking a promise herself, so far as her reasons go, as compared with others breaking promises. So the Act Consequentialist explanation of a constraint against promise-breaking is somewhat plausible. But the same kind of explanation may be less plausible for other constraints, such as a constraint against torture. Intuitively, the badness of torture is mostly a matter of the suffering caused to the victim, which is bad in an agent-neutral way. An Act Consequentialist explanation of a constraint against torture would have to claim, instead, that it would be wrong for me to torture someone when doing so would make the outcome better, in an impersonal way, because of the agent-relative badness of my act of torture, or of the relationship with my victim this would instantiate. In contrast, indirect theories need not appeal to agent-relative accounts of value, including the badness of torture, to explain constraints. They can say that the constraint against torture is explained by the agent-neutral badness of torture, which explains why a pattern involving the agent torturing someone is

disfavored compared to other eligible patterns. On the other hand, the challenge facing these theories is to develop a plausible account of the eligibility of patterns.

5. How do reasons interact?

A full theory of reasons would tell us several things about them. For example, it would tell us which reasons agents have, and why. It would also tell us how strong these reasons are. But a further question is how reasons interact with each other, and how their interaction relates to the rightness of actions. We touched on these issues in section 1, but it is worth briefly returning to them.

On a simple picture, an action is right if and only if there is no alternative action for which there is stronger reason. If this simple correlation holds, we would know which actions were right if we knew which reasons were strongest. If strength of reasons also *explains* rightness, as is somewhat plausible, we could identify which actions were right, and explain what makes them right, if we knew which reasons were strongest. This promise of explanatory elegance makes the simple picture attractive. But there are alternative possible pictures, with different merits. Earlier we postulated that the connection between rightness and reasons is that an action is right if there is "sufficient reason" to perform it. The simple picture says, in effect, that there is sufficient reason to perform an action just in case there is no alternative action for which there is a stronger reason. This is to assume that sufficiency of reason requires maximum strength of reason. This assumption is somewhat attractive, and it may seem to fit with the spirit of consequentialism.

However, nothing in the way we have defined consequentialism here mandates it. Consider briefly the merits of an alternative picture, according to which there is sufficient reason for an action just in case either there is no alternative action for which there is a stronger reason, or the agent has a moral right to perform the action. This alternative implies that it is never wrong to perform an action which you have a moral right to perform.[22] This would mean that, if consequentialists could account for moral rights, they would also be able to account for moral options. That is, they would be able to account for the idea that there is moral discretion in the exercise of moral rights, in the sense that it is not wrong for an agent

[22] More complex accounts of sufficient reason are also possible. For example, it is possible that it is usually right to perform an action for which you have a moral right, unless there is a *much* stronger reason to do something else.

to perform any of some set of alternative actions which would not have equally good outcomes. Since options (like constraints) are features of "common sense morality," it is an important question whether consequentialists are able to account for them.

As mentioned briefly in section 1, a further issue to consider is whether some kinds of reasons always or sometimes take priority over other kinds of reasons. For example, we might think that moral reasons sometimes but not always take precedence over other kinds of reason. If moral reasons are sometimes defeated by other reasons, what is right overall would diverge from what is morally right. In these and other ways, consequentialist theories of reasons can try to account for our convictions about the complexity of agents' reasons for action (Portmore, 2011, ch. 5; Sobel, 2007, 14–17).

6. Conclusion

The topic of consequentialist theories of reasons for action is under-explored when compared to the topic of consequentialist theories of the rightness of actions. Developing theories of reasons more fully would enable consequentialists to account for a greater part of what we care about in ethics. It would also create additional opportunities to introduce complexity into consequentialist ethical theories. Since common criticisms of consequentialism often involve the idea that it is too simple to account for the messiness of ethics, these opportunities are certainly worth exploring.

Developing consequentialist theories of reasons more fully should be of interest to non-consequentialists as well. It is often said that you do not have to be a consequentialist to think that the fact that an action would lead to a better outcome, for example, is (at least sometimes) a reason to perform it. Arguably, the same is true of the fact that an action would be your part in some beneficial or harmful pattern of action. Anyone who believes in either kind of reason has an interest in deepening our understanding of them.[23]

[23] I am very grateful to Douglas Portmore for very helpful comments on a draft of this chapter.

Bibliography

Alvarez, M. (2018). Reasons for action, acting for reasons, and rationality. *Synthese* 195, 3293–3310.

Bacharach, M. (1999). Interactive team reasoning: A contribution to the theory of co-operation. *Research in Economics* 53, 117–47.

Bardsley, N., Mehta, J., Starmer, C. & Sugden, R. (2010). Explaining focal points: cognitive hierarchy theory *versus* team reasoning. *The Economic Journal* 120, 40–79.

Bradley, B. (2018). Contemporary Consequentialist Theories of Virtue. In *The Oxford Handbook of Virtue*, edited by N. E. Snow, 398–412. Oxford: Oxford University Press.

Brandt, R. (1992). *Morality, Utilitarianism, and Rights.* Cambridge: Cambridge University Press.

Chappell, R. Y. (2015). Value Receptacles. *Noûs* 49 (2), 322–332.

Crisp, R. (2006). *Reasons and the Good.* Oxford: Clarendon Press.

Dietz, A. (2016). What We Together Ought to Do. *Ethics* 126, 955–82.

Foot, P. (1985). Utilitarianism and the Virtues. *Mind* 94 (374), 196–209.

Forcehimes, A. & Semrau, L. (2018). Are There Distinctively Moral Reasons? *Ethical Theory and Moral Practice* 21, 699–717.

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

Kagan, S. (1989). *The Limits of Morality.* Oxford: Clarendon Press.

Kagan, S. (2000). Evaluative Focal Points. In *Morality, Rules, and Consequences: A Critical Reader*, edited by B. Hooker, E. Mason, & D. Miller, 134–55. Edinburgh: Edinburgh University Press.

Lord, E. (2015). Acting for the Right Reasons, Abilities, and Obligation. In *Oxford Studies in Metaethics, Volume 10*, edited by R. Shafer-Landau, 26–52. Oxford: Oxford University Press.

Louise, J. (2004). Relativity of Value and the Consequentialist Umbrella. *The Philosophical Quarterly* 54 (217), 518–36.

Lyons, D. (1980). Utility as a Possible Ground of Rights. *Noûs* 14 (1), 17–28.

McElwee, B. (2010). The rights and wrongs of consequentialism. *Philosophical Studies* 151, 393–412.

Norcross, A. (1997). Good and Bad Actions. *Philosophical Review* 106 (1), 1–34.

Parfit, D. (1987). *Reasons and Persons.* Oxford: Clarendon Press.

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

Podgorski, A. (2018). Wouldn't it be Nice? Moral Rules and Distant Worlds. *Noûs* 52 (2), 279–294.

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

Portmore, D. W. (2019). *Opting for the Best: Oughts and Options.* Oxford: Oxford University Press.

Ridge, M. (2006). Introducing Variable Rate Rule Utilitarianism. *The Philosophical Quarterly* 56 (223), 242–53.

Roberts, M. A. & Wasserman, D. T. eds. (2009). *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem.* Dordrecht: Springer.

Ross, W. D. (2002/1930). *The Right and the Good*, ed. P. Stratton-Lake. Oxford: Clarendon Press.

Scanlon, T. M. (1998). *What We Owe to Each Other.* Cambridge, MA: Belknap Press.

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

Scheffler, S. (1992). *Human Morality*. New York: Oxford University Press.

Schroeder, M. (2007). Teleology, Agent-Relative Value, and "Good." *Ethics* 117, 265–295.

Smith, H. M. (2010). Measuring the Consequences of Rules. *Utilitas* 22 (4), 413–33.

Smith, H. M. Alternatives. This volume.

Snedegar, J. (2017). *Contrastive Reasons*. Oxford: Oxford University Press.

Sobel, D. (2007). The Impotence of the Demandingness Objection. *Philosophers' Imprint* 7 (8), 1–17.

Star, D. ed. (2018). *The Oxford Handbook of Reasons and Normativity.* Oxford: Oxford University Press.

Ullmann-Margalit, E. & Morgenbesser, S. (1977). Picking and Choosing. *Social Research* 44 (4), 757–85.

Way, J. & Whiting, Daniel. (2017). Perspectivism and the Argument from Guidance. *Ethical Theory and Moral Practice* 20, 361–374.

Woodard, C. (2009). What's Wrong with Possibilism. *Analysis* 69 (2), 219–26.

Woodard, C. (2013). The Common Structure of Kantianism and Act Utilitarianism. *Utilitas* 25 (2), 246–65.

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