Philosophia (2018) 46:207–221
DOI 10.1007/s11406-017-9908-y

Does 'Ought' Imply 'Might'? How (not) to Resolve the Conflict between Act and Motive Utilitarianism

James Skidmore[1]

Received: 14 June 2016 / Accepted: 12 September 2017 /
Published online: 26 September 2017

Abstract Utilitarianism has often been understood as a theory that concerns itself first and foremost with the rightness of actions; but many other things (e.g., moral rules, motives, laws, etc.) are also properly subject to moral evaluation, and utilitarians have long understood that the theory must be able to provide an account of these as well. In a landmark article from 1976, Robert Adams argues that traditional act utilitarianism faces a particular problem in this regard. He argues that a on a sensible utilitarian account of the rightness of an agent's motives, right motives will sometimes conflict with right actions, leaving the theory internally incoherent. The puzzle Adams raises has received a good deal of attention but few proposed solutions. Fred Feldman, however, has offered a solution that seems to be gaining adherents. In this paper I argue that Feldman's approach cannot succeed. At bottom, it relies on a version of the principle that 'ought' implies 'can'—and subsequently an account of an agent's alternatives—that is far too restrictive to be plausible. Despite the failure of this solution, however, I argue that the conflict Adams develops is not as theoretically troubling as he suggests. While traditional act utilitarianism may fail for other reasons, it will not fail due to the conflict between acts and motives.

Keywords Utilitarianism · Consequentialism · Motive · Ought implies can · Blameless wrongdoing

1 Introduction

Utilitarianism has often been understood as a theory that concerns itself first and foremost with the rightness of actions. Thus at the heart of traditional act utilitarianism

✉ James Skidmore
skidjame@isu.edu

[1] Department of English and Philosophy, Idaho State University, Campus Box 8056, Pocatello, ID 83209-8056, USA

is the account of right action. Roughly, an act is right if and only if it maximizes overall, long-run happiness relative to alternatives.[1] But acts are not the only things that are subject to moral evaluation, and utilitarians have long understood that, in order to be complete, utilitarianism must provide an account of the moral evaluation of everything that is properly subject to such evaluation.

In a landmark article, Robert Adams (1976) argues that traditional act utilitarianism faces a particular problem in this regard. He argues that on a sensible utilitarian account of the rightness of an agent's *motives*, right motives will sometimes conflict with right actions, leaving the theory internally incoherent. The puzzle Adams raises has received a good deal of attention but few proposed solutions. Fred Feldman, however, has offered a solution that seems to be gaining adherents. Feldman develops an account of utilitarianism—and implicitly an account of an agent's *alternatives*—that dissolves the alleged conflict and ensures harmony between right motives and right actions.

In what follows I argue that Feldman's approach cannot succeed. At bottom, it relies on a version of the principle that 'ought' implies 'can'—and subsequently an account of an agent's alternatives—that is far too restrictive to be plausible. Despite the failure of this solution, however, I argue that the conflict Adams develops is not as *theoretically* troubling as he suggests. While traditional act utilitarianism may fail for other reasons, it will not fail due to the conflict between acts and motives.

2 The Conflict: the Cases of Jack and Clare

Adams argues that the motives of an agent—the agent's desires or "patterns of motivation"—should be an important target of moral evaluation. How, then, should a utilitarian evaluate them? He argues that the most plausible utilitarian account would have us evaluate an agent's motives, just as we evaluate her acts, according to their utility. "[O]ne pattern of motivation is morally better than another to the extent that the former has more utility than the latter. The morally perfect person, on this view, would have the most useful desires, and have them in exactly the most useful strengths" (Adams 1976: 470).

Adams calls this "motive utilitarianism." On this view, the rightness of an agent's motives will be determined by calculating their utility, giving us something like the following definition:

> A motive (or motivational pattern) is right if and only if it maximizes overall, long-run happiness relative to alternatives.

On Adams' view, then, motive utilitarianism runs parallel to act utilitarianism, and the two are theoretically distinct. One might assume, however, that a comprehensive utilitarian theory would accept both claims, appealing to utility-maximization to account not only for the rightness of actions but also of motives and other targets of moral evaluation (character, conscience, etc.). This generates direct utilitarianism: *for any x*

[1] There are of course alternative utilitarian accounts of the good. As far as I can tell, nothing in this paper turns on which version is chosen. Indeed, nothing in the paper turns on whether utilitarianism or some other consequentialist theory is chosen.

(where x is properly subject to moral evaluation), x is right if and only if it maximizes overall, long-run happiness relative to alternatives. (See, e.g., Kagan 2000; Pettit and Smith 2000; Parfit 2011: 374).

Adams argues, however, that act and motive utilitarianism (as outlined above) will conflict in practice in the following way: sometimes the right motive for an agent to have will lead the agent to perform an act that is wrong by act-utilitarian standards. To illustrate, Adams gives us the case of Jack. Jack is making his first visit to the cathedral at Chartres, and he is enthralled. In his enthusiasm, he does his best to take in the cathedral in all its detail. This enthusiasm is leading him to stay much longer than he expected, examining even the less interesting and enjoyable parts of the cathedral. Just now, for example he is "studying the sixteenth to eighteenth century sculpture on the stone choir screen" (Adams 1976: 471). He is enjoying this less than other parts of the cathedral, and will not remember it very well. This focus in turn leads to "considerable inconvenience and unpleasantness," as he remains so long that he misses dinner and must drive several hours that night to find a room. In fact, the time Jack spent on the least enjoyable parts of the cathedral was inefficient—he would have maximized utility by skipping the choir screen and other less interesting parts and leaving earlier. Jack realizes this while he is in the cathedral, but he doesn't much care. He wants to see the cathedral in all its detail, and has no great concern to "squeeze out the last drop of utility" (Adams 1976: 471).

According to act utilitarianism, then, Jack fails to do what is right. The right course of action would have been to focus on the most interesting parts of the cathedral and leave earlier. Adams suggests, however, that Jack's *motive*—his devotion to seeing as much of the cathedral as possible—was the *right* one to have by motive-utilitarian standards, because it led him to enjoy his visit more intensely than he would have had he toured the museum with a different motive (say, the motive of maximizing utility). Here we have a situation, then, in which the agent's motive is right by motive-utilitarian standards: Jack's motive of seeing as much of the cathedral as he can (utility be damned) is the right one for him to have; it maximizes utility relative to alternative possible motives. However, this motive leads him to perform actions that are wrong by act-utilitarian standards; his enthusiasm leads him to linger even at the less interesting parts when the alternative of passing them over and leaving earlier would have been better. The right motive leads to the wrong action; motive and act utilitarianism conflict.[2]

Derek Parfit (1984) examines the same kind of conflict in developing the notion of "blameless wrongdoing." He offers us the case of Clare. Clare is a confirmed consequentialist, and she is also a mother. Clare understands, however, as Adams did, the

[2] Feldman develops a helpful interpretation of the case, one that makes the conflict vivid. First, he labels the relevant motives and actions as follows:

"Mmax: The motive of wanting to maximize utility.
Msee: The motive of wanting to see as much as possible at the cathedral.
A1: The act of studying the cathedral during the morning with motive Mmax.
A2: The act of studying the cathedral during the morning with motive Msee.
A3: The act of leaving early in the afternoon.
A4: The act of staying in the cathedral in the afternoon."If we assign A1 a utility of 5; A2 a utility of 10; A3 a utility of 2; and A4 a utility of -2, then the conflict becomes apparent. MSee is the right motive (with a utility of 8, one more than MMax's 7), but it leads Jack to perform A4, which is the wrong act. See Feldman (1993: 205).

moral importance of having good motives. She realizes that "[m] ost of the best possible sets of motives would include strong love for our children" (Parfit 1984: 32). Thus Clare has such a love for her child as a part of the best set of motives she can have. But now suppose that Clare finds herself in a situation in which she "could either give her child some benefit, or give much greater benefits to some unfortunate stranger" (Parfit 1984: 32). According to act consequentialism, the right act in such a case is to benefit the stranger; but Clare loves her child, and this leads her to benefit her child instead. Here again we have a case of what Adams sees as a conflict between act and motive utilitarianism: Clare has the right motives according to motive utilitarianism, but they lead her to act wrongly in this case. Parfit does not discuss the case as one of theoretical conflict; he instead sees it as a case of blameless wrongdoing. It is wrongdoing because, according to act consequentialism, she does the wrong thing, but blameless because there is no reason for her to feel guilty for, or for others to resent, what she has done. After all, such guilt and resentment could only undermine the love she feels for her child, and by hypothesis that love is optimal. Here again, then, we have a case in which the right consequentialist motives lead to the wrong consequentialist act.

3 Resolving the Conflict

While these sorts of cases have generated a good deal of discussion among proponents and critics of consequentialism, there has been no clear agreement on how to address them. In recent years, however, there is one particularly noteworthy strategy that seems to be gaining support among (direct) utilitarians. It is a strategy that eliminates any appearance of conflict in such cases by implicit appeal to a distinct interpretation of an agent's alternatives and the underlying principle that 'ought' implies 'can'. Feldman (1993) provides perhaps the earliest example of this approach in his proposed resolution of Adams' original case. Feldman acknowledges that Adams has identified a conflict between act and motive utilitarianism as he (Adams) formulates them, but Feldman argues that on his own version of utilitarianism the conflict dissolves. Feldman's version is a kind of "possible world" utilitarianism. He begins with the notion of an "accessible world." At any particular time, there will be a variety of possible worlds accessible to an agent—"various possible ways in which she might live out her life" (Feldman 1993: 208). These worlds can be ranked according to the overall utility they contain. This leads to a basic moral principle. For any agent s at time t:

> "U: As of t, s morally ought to see to the occurrence of p iff p is true in all the bests [best accessible worlds] for s at t" (Feldman 1993: 209).

In other words, what an agent ought to do is to see to the realization of the best accessible world, and this applies not just to acts to but to anything else as well: her motives, character, conscience, anything relevant to realizing the best world available to her.

Feldman then applies this version of utilitarianism to the case of Jack. When Jack awakes that fateful morning, the best world accessible to him is one in which he is motivated by the desire to see the entire cathedral ("*Msee*") (*not* the desire to maximize utility, "*Mmax*") (Feldman 1993: 205). This motive, of course, will lead him to stay late

at the cathedral, miss his dinner, etc. All of these things are part of any world in which he has *Msee*, and thus they are *all* part of the best possible world accessible to him; therefore, on Feldman's view, they are all things that he should do or see to. But if this is true, then the supposed conflict has disappeared: when Jack stays late at the cathedral, he is doing the right thing! Both his motives and his acts are the right ones because they are all part of his best accessible world.

Feldman sees only one way of avoiding this result: "Perhaps Adams intended that it would be possible at 2:00 pm for Jack to leave early even if he had adopted *Msee* in the morning" (Feldman 1993: 211). If this is true, Feldman argues, then there was another world accessible to Jack that morning, one in which he views the cathedral with one motive, then changes to the motive of maximizing utility later in the day in order to leave early. But if this world is accessible to him, then this is the world he should see to. In either case, there is no conflict between the motives he ought to have and the actions he ought to perform.

Jonathan Dancy adopts a very similar strategy to dispel the alleged conflict in Parfit's case of Clare. He begins by emphasizing that Clare acts wrongly in benefiting her child only if there is a better alternative available, and he goes on to suggest that there is not. Parfit claims that the alternative of benefiting the stranger is better, but Dancy evaluates this claim by asking us to examine the nearest possible worlds in which Clare chooses this option. In these worlds, he claims, Clare does not possess the optimal set of motives, and her sub-optimal motives lead to a worse overall outcome. "So the nearest group of worlds in which she [benefits the stranger] is a group in which the outcomes are worse in C's terms" (Dancy 1997: 15).

While both Feldman and Dancy attempt to resolve the alleged conflict in these cases by developing a 'possible worlds' analysis of them, the crux of each solution is a novel account of an agent's alternatives, together with an implicit appeal to the principle that 'ought' implies 'can'. Central to direct utilitarianism is the idea of an agent choosing the best act among alternatives. But what counts as an alternative? Feldman's and Dancy's solutions manage to eliminate any conflict between right motive and right act in these cases by eliminating one (allegedly optimal) act as an alternative.

Feldman's discussion makes clear that, on his view, one of Jack's apparent alternatives is no alternative at all: that is the alternative of leaving the cathedral early *despite Msee*, his desire to see the whole thing. For Feldman, there is no accessible world in which Jack *keeps Msee* yet still leaves the cathedral early; as he says, "the choice of *Msee* in the morning rules out all possible worlds in which he leaves the cathedral early" (Feldman 1993: 211). In this way, the conflict is dissolved by narrowing Jack's alternatives: Given that the right motive for Jack to have is *Msee*, and supposing it is not something he can simply change at a moment's notice, the alternative of leaving the cathedral early is no alternative at all. Jack *cannot* leave the cathedral early because he has *Msee*. Of course, if leaving early is no alternative, then it cannot be wrong to stay late—we can only expect Jack to perform the best acts among the alternatives he has.

Dancy's resolution of Parfit's case works in exactly the same way. Dancy suggests that Clare's alternative of benefiting the stranger is not really optimal, because in the nearest worlds in which she performs this act she has a worse set of motives, resulting in an overall worse outcome. But this analysis can only succeed in resolving the conflict by eliminating one apparently available alternative: Clare's benefiting the stranger *despite* her optimal motive set. Dancy must claim here that if we hold Clare's

optimal set of motives fixed, then benefiting the stranger is no alternative; for if it *is* an alternative, then by hypothesis it is the optimal one. Just as Feldman must hold that Jack *cannot* leave early given his optimal motive of *Msee*, so Dancy must hold that Clare *cannot* benefit the stranger given her optimal motive set. In each case, what appeared to be the right act turns out be an act that the agent *cannot* perform given his or her optimal motive, and thus it is not an alternative at all.

Elinor Mason has recently sided with Dancy in his resolution to Parfit's case of Clare, and her account makes more explicit this fundamental reliance on narrowing Clare's alternatives. In describing Clare's situation, Parfit suggests that her optimal love for her child rules out the possibility that she will benefit the stranger; benefiting the stranger is something that she would do *only* if she loved her child less. But Mason argues that if this is true, then the "alternative" of benefiting the stranger, given her love, is not an available alternative at all; it is not "causally possible" that she benefits the stranger while loving her child optimally (Mason 2002: 290). Since consequentialism only demands that we choose the best among the available alternatives, Clare does nothing wrong in benefiting her child instead of the stranger. In developing her optimal love for her child, Clare has "tied herself to the mast", and benefiting the stranger is something she *cannot* do. The right motive no longer conflicts with the right act.

While Feldman, Dancy, and Mason thus develop their accounts in different ways, in the end they make use of the same basic strategy. It is a strategy that focuses on an agent's alternatives and an implicit appeal to the principle that 'ought' implies 'can'. It simply applies this principle in a way that takes the motives of the agent into consideration. First, since 'ought' implies 'can', and since right utilitarian acts are only those that are best among alternatives, an agent's act cannot be wrong unless there is some better act that it is possible for him to perform. Second, an agent's motives may well rule out the possibility of his performing certain acts: Given a certain motive, it may simply not be possible that an agent will perform a particular act. Thus, given Jack's and Clare's motives, we might suppose that it is simply not possible that he will leave the cathedral early or that she will benefit the stranger. But if their motivations rule these alternatives out, then Clare and Jack have not acted wrongly after all. Given their (optimal) motives, they have done the best they can do. In each case, it is argued that what appear to be the optimal acts are not among the agent's alternatives at all precisely because they are incompatible with the agent's optimal motive.

This basic approach to the resolution of conflict between act and motive utilitarianism, and to the dissolution of blameless wrongdoing, has gained other recent adherents as well. Bart Streumer joins Mason and Dancy in dismissing Parfit's case of Clare with this strategy, repeatedly insisting that "Clare cannot both love her child and benefit the stranger" (Streumer 2003: 243). Others have defended the essential claims of the approach outside the context of these particular problem cases. For example, a number of authors have endorsed the notion that an agent's motives can render her unable to perform an act that she would otherwise be perfectly able to perform (so that, because 'ought' implies 'can', it must be false that she ought to perform that act, given that motive) (See Crisp 1992; Bloomfield 2001: 172; Streumer 2007; Anomaly 2008). Despite this growing support, however, I argue in the next section that the approach cannot succeed.

4 Problems

For convenience, let us call this basic approach the Argument from Motivational Incapacity. It is a tempting line of reasoning. Both of the main claims it relies on seem initially plausible. First, it seems absurd to claim that an agent ought to have performed an act that he *could not* have performed; and second, it seems plausible to think that our motives do effectively rule out certain acts, rendering them (in some ultimate sense) impossible. Yet if an act is impossible–if it *cannot* happen–how can it be a genuine alternative?

Despite this intuitive appeal, we may sense that something is amiss. Notice that in attempting to resolve any conflict in the cases of Jack and Clare, Feldman and Dancy have essentially *redescribed* the cases. What began as ordinary cases in which an agent's motives led her not to perform a particular act have now been redescribed as extreme cases in which an agent's motives left her *unable* to perform a certain act. While it seems obvious that our motives routinely have an influence on our actions, in ordinary cases we do not conclude that they leave us literally *unable* to perform acts that we would otherwise be able to perform. We may well make exceptions in certain cases—admitting, say, that someone's addiction to heroine leaves him genuinely unable to quit, or that someone's intense fear of heights could leave her literally unable to walk out to the cliff's edge. But in ordinary cases our motives lead us away from a particular course of action without leaving us unable to perform it. My modest preference for coffee, for example, will normally lead me to choose it over tea without leaving me *unable* to choose the latter.

Originally the cases of Jack and Clare were intended to be cases of the ordinary kind, cases in which an agent's optimal motive led her not to perform an optimal act while leaving her *able* to perform that act. In describing Clare's case, for example, Parfit explicitly stipulates that her act is voluntary: "She could, if she wanted, avoid doing what she believes to be wrong. She fails to do so simply because she wants to benefit her child more than she wants to avoid wrong-doing" (Parfit 1984: 32). In other words, Clare's optimal motive does *lead* her to perform the wrong act, but not by leaving her *unable* to perform the optimal act.

It seems plausible to interpret Jack's case in the same way. It must be admitted that Adams' description of the case sometimes wavers. He says at one point that Jack, given his motive of seeing the entire cathedral, "could not bring himself to leave the choir screen as quickly as would have maximized utility" (Adams 1976: 472). This description encourages the interpretation that Jack's motive is so overwhelming that it renders him incapable of leaving, so that even if he decided to leave he "could not bring himself" to do so. Perhaps Feldman is interpreting the case in this way when he suggests that, given Jack's motive, it is not within his power to leave the cathedral early, and so that leaving early is not a genuine alternative (Feldman 1993: 211). But this does not appear to be what Adams has in mind. He insists later that it *is* in Jack's power to leave the choir screen "if he wants to; it is just that he does not want to" (Adams 1976: 473). Again, Jack's optimal motive *leads* him to perform the wrong act, but not by leaving him *unable* to perform the right one. On Feldman's and Dancy's readings, however, the cases have been redescribed;

they are now cases in which Jack's and Clare's optimal motives leave them unable to perform the allegedly optimal act.

The source of this impasse is not obvious, but it appears to lie in a divergence in the crucial implicit assumptions made by each side in interpreting these cases. On reflection, we can see that Adams and Parfit rely on a few key assumptions in developing the cases of Jack and Clare. First, their discussion assumes that there are limits on an agent's ability to change her motives. Motives, i.e., desires, commitments, traits of character, etc., at least in some cases, are such that an agent cannot—*is not able to*—simply adopt them and drop them at will, instantaneously. If this *were* possible—i.e., if Clare were able to "drop" her optimal love for her child at a moment's notice, then "pick it back up" immediately later on—then again the alleged conflict dissolves; for now, *at the moment of choice*, Clare's love for her child is not the optimal motive. She should instead drop her love momentarily and adopt a motive that will lead her to help the stranger. To put it another way, if motives were like this, then the optimal set of motives would *change* constantly according to the situation. Adams and Parfit must reject this possibility, and Parfit does so explicitly. As he says, "it is in fact impossible that our love could be like this. We could not bring about such 'fine-tuning'" (Parfit 1984: 34). The conflict cases arise only because the agent has a particular motive at a particular time, and *cannot change it* at that moment.

Second, their discussions clearly assume that motives have an influence on human action. If this were not true, then there could be no conflict between optimal motive and optimal act in these cases. For example, if Jack's motive has no influence on his decision not to leave the Cathedral at the optimal time, then it cannot be said to conflict in any way with that act.

But these cases also assume something further about the nature of this influence. They assume that a sort of determinism often holds between motives and acts. Parfit calls this "Psychological Determinism," the view that "our acts are always caused by our desires, beliefs, and other dispositions" (Parfit 1984: 14).[3] On this view, Clare's optimal motive set causally determines the act she will perform—helping her child instead of the stranger, just as Jack's motive *MSee* guarantees his decision to remain at the cathedral. In each case, the agent's optimal motive determines the act she will perform.

While this sort of motive-act determinism may well be plausible, it should be noted that these cases actually rely only on a weaker, negative claim. In order to generate the relevant conflict between optimal motive and optimal act, we need only assume that these are cases in which the optimal motive determines that the optimal act will *not* be performed. That is, we assume only a negative motive-act determinism, according to which motives sometimes have the effect of ruling certain actions out, of making it inevitable that they will not be performed. For Jack, the optimal act in the afternoon is to leave the cathedral, but his motive makes it inevitable that he will not do this (whether or not it determines exactly what he *will* do instead). Clare's optimal motive similarly guarantees that she will not benefit the stranger. These cases, then, rely not only on the assumption that motives influence actions, but also on a weak, negative

[3] Parfit later considers the possibility that "Psychological Determinism" is false, and argues that the conflict between act and motive in Clare's case remains. I discuss this alternative below.

form of motive-act determinism: sometimes our motives determine that an act will not be performed.

While Adams and Parfit do rely on these assumptions in interpreting the cases of Jack and Clare, it is hard to imagine them being controversial. Intuitively, we do think that our motives are to a significant degree subject to our voluntary control. But surely Parfit is correct, for example, in claiming that a parent cannot simply drop her strong love for her child at a moment's notice. This is something she could only do over time. The claim that our motives determine inevitably that some acts will not be performed may seem more controversial, but on reflection it seems just as obviously true. An agent's food preferences, for example, will routinely have the effect of ruling out certain courses of action; my modest dislike of Brussells sprouts, together with the availability of several more palatable alternatives, *does* effectively rule out the possibility that I'll be eating them for dinner. I simply won't, because I don't like them.

Given the intuitive plausibility of these assumptions, it is not surprising that critics such as Feldman and Dancy seem to accept them.[4] Instead, what underlies their reinterpretation of these cases is the rejection of a further assumption on which Adams and Parfit depend. I have suggested above that these cases rely on a weak form of motive-act determinism, namely, the claim that an agent's motives routinely determine that certain acts will *not* be performed. Regarding this weak form of determinism, both Adams and Parfit assume compatibilism. That is, they assume that this determinism is compatible with the agent's *ability* to perform these acts. So, for example, Adams assumes that, while Jack's motives *guarantee* that he will not leave the cathedral early, it still remains true that he is *able* to leave early. Similarly, Parfit assumes that Clare remains *able* to help the stranger, even though her optimal motives determine that she will not. In short, both Adams and Parfit assume that my ability to perform a particular act is compatible with that act being ruled out by the motives I possess.[5] There are some acts that I am able to perform, but which, because of my motives, I *will* not.

It is *this* assumption that Feldman and Dancy both implicitly reject in reinterpreting these cases. They adopt an *incompatibilist* approach to the weak form of motive-act determinism on which the cases rely.[6] As we saw above, both authors essentially argue that because Jack's and Clare's motives determine that the alleged optimal act will not be performed, together with the fact that they cannot at the moment change those motives, these are acts that they simply cannot perform. They assume that an agent's ability to perform a particular act is *not* compatible with having a set of motives which rule the act out. Thus, on their reinterpretation, the alleged optimal acts are not really alternatives at all, and it cannot be the case that the agents ought to perform them. In this way, we see that the source of this dispute lies in a deeper disagreement between a compatibilist and an incompatibilist approach to the weak motive-act determinism assumed in the cases.

[4] Louise (2006) seems to criticize them for assuming that motives causally determine actions, but here they are simply adopting the assumption originally made by Adams and Parfit. Dancy and Mason do go on to raise questions with Parfit's assumption of "psychological determinism." As noted above, Parfit argues that even if such determinism is false the conflict between act and motive in Clare's case remains. I discuss this alternative below.

[5] Parfit explicitly defends this form of compatibilism in his more recent work. See Parfit (2011: 260).

[6] It is interesting to note that Louise, despite *opposing* Feldman and Dancy's rejection of the conflict cases, seems to follow them here in assuming such incompatibilism. See Louise (2006: 79).

How then are we to adjudicate this compatibilist/incompatibilist dispute? It must first be emphasized again that the version of determinism in question here is not the one that ordinarily concerns philosophers engaged in debates regarding the compatibility of free will with determinism. In those debates, determinism is "the thesis that a complete description of the state of the world at any time t and a complete statement of the laws of nature together entail every truth about the world at every time later than t" (Vihvelin 2015). Whether an agent's freedom (and thus his ability to perform alternative actions) is compatible with *this* form of determinism is a matter of long-running dispute that I cannot hope to settle here. We are considering instead a much weaker form of "determinism," the claim that an agent's motives sometimes determine that certain acts will not be performed.

In fact this version of determinism is sufficiently weak that its truth, together with Adams' and Parfit's compatibilist approach to its truth, seem intuitively obvious. It seems obvious, first, that my motives at any particular time *do* effectively rule out the possibility of my performing certain actions. My current set of motives, for example, does not include any desire to see the movies now playing at nearby cinemas. Because of this, I *will not* be going to the cinema in the next few days. I simply won't, because I don't want to. Similarly, a compatibilist approach to this weak form of determinism seems equally obvious. Surely it would be bizarre to claim that my lack of interest in these current movies leaves me *unable* to see them. My ability to see them is not in question (as it would be, for example, if I had no money), but rather my *willingness*. On reflection, it seems obvious that there is, at any one time, a long list of actions that my current motives leave me unwilling to perform. It is implausible to say, of *all* of these, that I am unable to perform them. Yet that is what Feldman's and Dancy's incompatibilism implies.

The implausibility of this incompatibilism can be seen as well in the highly unusual interpretation that it generates of the principle that 'ought' implies 'can'. On this view, any act that an agent will not perform, given her current motives, is one that in fact she *cannot* perform. Thus, on this view, an agent is able to perform a particular act only if, given her current motives, she might actually perform it; 'ought' implies 'might'!

But clearly something has gone wrong, for 'ought' does not imply 'might' in this way. In claiming that an agent A ought to Φ, I am *not* committed to the claim that he might in fact Φ. For example, my sincere belief that Henry Kissinger ought to turn himself in to the Hague to be prosecuted for war crimes does not commit me to believing that he *might* in fact do this. The contrapositive is even more obvious: The claim that there is no chance that A will Φ does not imply that it is not the case that A *ought* to Φ. For example, there may be no chance that I will donate more than 20 % of my income this year to charitable organizations fighting poverty; perhaps I am simply too stingy to do that. This truth does not by itself imply that it is false that I ought to do so.

The intuitively strange implications of this commitment to the claim that 'ought' implies 'might' can be seen nicely in an example developed by Mason. She considers the hypothetical case of Angela, who is "a very malevolent and lazy person. She could try very hard to improve her motives, but she would fail, and the results would be very bad." Mason suggests then that Angela's motives, as bad as they are compared with those of other agents, may actually be the best motives she can have (Mason 2002: 295).

Mason accepts the consequentialist implication: In Angela's case, her laziness and malevolence really are the right motives for her to have. But let us now suppose further that 'ought' implies 'might', so that any act that is incompatible with Angela's laziness and malevolence is an act that she is *unable* to perform. One morning Angela is, as usual, driving her large SUV three blocks to work. As she passes by the small park with the wading pool, she sees a toddler face down in the water, struggling feebly. What should she do? Intuitively, she ought to rescue the child; but suppose that this alternative turns out not to be consistent with her motives: given her malevolence and laziness, she simply *will not* make any effort to save the child. If 'ought' implies 'might', then what appears to be the optimal alternative—rescuing the child—is not an alternative at all. Angela continues driving, the child drowns, and we are forced to conclude that she did the right thing!

In short, Feldman, Dancy, and Mason are able to eliminate the alleged conflict in the cases of Jack and Clare only by reinterpreting them in such a way that the alleged optimal act that conflicts with the optimal motive—for Jack, the act of leaving early; for Clare, the act of benefiting the stranger—is one that they *cannot* perform. This is accomplished by adopting an incompatibilist approach to the negative form of motivational determinism on which these cases depend. But this incompatibilism has the effect of shrinking implausibly the class of actions we are capable of performing at any one time. In particular, it essentially eliminates one subset of this class: the set of acts that we are able *but unwilling* to perform. It seems obvious that, at any one time and for any particular agent, there are a vast number of acts in this set—acts that the agent certainly could perform, but which, given her motives, she will not.

It seems unlikely, then, that the conflict in these cases can be resolved by taking an incompatibilist approach to the negative motivational determinism on which they depend. Perhaps it is the determinism itself that should be rejected. Mason, for example, argues that if the 'psychological determinism' that Parfit assumes is false, then the case of Clare presents no conflict at all. "If Clare genuinely could have benefited the stranger, (in a completely straightforward sense of 'could have') and doing so would not have turned out worse in the long run (by weakening her optimal motives for example), then a consistent consequentialist must say that she ought to benefit the stranger" (Mason 2002: 289). Bart Gruzalski defends a similar view in an earlier discussion of Parfit's case. He claims that if Clare could have benefited the stranger without altering her optimal motive set, then her failure to do so "is wrong, she was morally bad, and there is no divergence between MU and AU" (Gruzalski 1986: 773).

Parfit agrees, of course, that Clare ought to have benefited the stranger and that her failure to do so is wrong. The crux of Gruzalski's and Mason's response here is that there is no conflict between Clare's optimal motive set and the act of benefiting the stranger as long as her motives leave her *capable* of performing that act. But this is a very narrow interpretation of what would constitute conflict between an act and a motive set. Suppose that Clare's optimal motive set does not completely rule out the possibility that she will benefit the stranger, but that it strongly inclines her not to do so and to benefit her child instead. Suppose that if she does decide to benefit the stranger, it will be *despite* her optimal motive set—in particular, despite her love for her child.

Suppose that she ultimately decides not to benefit the stranger, though she might have done so, and she decides this *because* of her motives—i.e., she decides to benefit her child instead *because* of her (optimal) love for the child.

Surely it is plausible to say that her motive set *conflicts* with the optimal act in this case. While her motives did not rule that act out entirely, they strongly inclined her not to perform it and ultimately constituted the *reason* she did not. In short, we now have a case in which an agent's optimal set of motives strongly incline her *not* to perform the optimal act, a case in which the optimal set of motives does not encourage, but rather *discourages* the agent from doing the right thing.

We see then that Adam's and Parfit's cases of conflict can survive the rejection of even the weak form of motivational determinism on which they originally relied. The attempts by Feldman, Dancy, and others to redescribe the cases so as to eliminate all conflict do not succeed. We are left to face a fundamental disharmony within a direct-consequentialist account of motives and acts: Sometimes an agent's morally right set of motives will lead him to perform the morally wrong act.

5 Return to the Conflict

Has Adams then succeeded in exposing a conflict at the heart of utilitarian theory? It is not clear that he has. At the end of his paper, Adams acknowledges that one might accept his conclusion that right motives sometimes lead to wrong actions, but still deny that there is any theoretical conflict. We simply admit that a person with the best motives (and the best conscience, the best character, etc.) will sometimes do what is wrong. There is no *theoretical* conflict here unless we suppose that somehow the right motives *must* by definition always produce right actions. But why suppose that? The primary goal of a moral agent, we might argue, is to live a moral *life*, which according to utilitarianism would involve trying to maximize utility over the course of a lifetime.[7] The primary goal is not to perform as many right acts as possible. One could easily imagine a life that contains *no* optimal acts, yet in the long run produces more utility than an alternative life that contains many optimal acts.[8]

Of course, it would be absurd (and not utility-maximizing) to *blame* an agent for wrong acts that result from right motives, and so the utilitarian argues that these acts will constitute a sphere of blameless wrongdoing (as we have seen Parfit defend above). This is not a new idea; it is one version of the familiar point that (act) utilitarianism makes a sharp distinction between the wrongness of acts and the blameworthiness of agents.

Adams understands that an act utilitarian can in this way elude the theoretical conflict, but he thinks there is a price to pay: Such a sphere of "blameless wrongdoing," he argues, has the effect of trivializing the notion of moral obligation. Common sense suggests that moral obligations or moral wrongs are serious matters; a morally good person will at least *try* to fulfill her obligations and do what is right, and she will care

[7] See, for example, Railton's (1984) development of "sophisticated consequentialism."

[8] Consider an analogy: The primary goal of a professional tennis player is to win matches; it is *not* primarily to hit has many optimal shots as possible, or even worse, never to hit a bad shot.

when she fails.[9] Act utilitarianism, in avoiding the conflict Adams raises, now claims that this isn't necessarily so: sometimes a morally good person will knowingly do something morally wrong and not care. If, on the other hand, part of what we *mean* in calling an act morally wrong is that it is blameworthy, then this act-utilitarian account will have failed to capture the meaning of the term.[10]

The defender of utilitarianism has at this point a choice. Shall the theory 1) follow act utilitarianism and define wrong action directly in terms of utility, thus separating "wrongness" sharply from blameworthiness and creating a sphere of acts that are morally wrong—but trivially so; or 2) depart from act utilitarianism and define right actions as those that are in harmony with the best motives (or conscience, or character), thus *linking* wrongness and blameworthiness, while creating a realm of sub-optimal acts (perhaps even *disastrous* acts) that are not wrong?

I will not try to settle here which overall course is the wisest. Despite Adams' concerns, however, I do not think that the former course—that of the traditional act-utilitarian—is obviously untenable. Does the existence of blameless wrongdoing threaten to trivialize the notion of moral wrong? It seems to me that it need not. To see why, we might consider an analogy.

Just as a conscientious moral agent will be properly concerned with the rightness or wrongness of her actions, so will a *rational* person possess a deep concern for the truth of her beliefs. But what form will this concern take? How far will it extend? Will a conscientious rational agent be profoundly concerned with the truth of *all* her beliefs? Every *single* one? Or could she instead be indifferent to the fact that some of her beliefs are (trivially) false and yet still qualify as rationally blameless, thus establishing a sphere of "blamelessly false beliefs?"

Consider once again the case of Jack. During his tour of the cathedral at Chartres, he hears a tour leader discussing the fire that burned the cathedral in 1194. This catches his attention, as he seems for some reason to remember reading that the fire occurred in 1195. Realizing that he is often forgetful about such details, he briefly considers seeking out pen and paper so that he can make a note of this fact. He decides against it. It is too much trouble, he thinks, and anyway it doesn't really matter. He moves along. Two years later, back at home, he has many fond memories of the cathedral but has indeed forgotten the correct date of the fire. Showing his pictures to a friend, he mentions the late twelfth-century fire—1195, he seems to remember—that destroyed the church soon after it was first built….

Jack has a false belief. What is worse, he had the perfect opportunity to correct it and *purposely failed* to do so. Yet worse, as much as he loves the cathedral, he doesn't even *care* much about whether this belief is (exactly) true. Shall we conclude that Jack is rationally blameworthy for such indifference to the truth? We might conclude instead that there is a legitimate sphere of "blamelessly false belief," and that it is compatible

[9] This concern may also help explain Adams' choice of the example of Jack in the Cathedral to illustrate the conflict he develops. We wouldn't normally think that a decision between spending the afternoon in the cathedral and finding a room is a *moral* decision at all. The fact that act utilitarianism implies that Jack does something *morally* wrong in staying late and condemning himself to an uncomfortable night illustrates by itself the way in which Adams thinks the theory has trivialized the concept of moral obligation.

[10] Mill famously suggests this view in Chapter 5 of *Utilitarianism*: "We do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience" (Mill 1979: 93).

with the more general and profound concern for the truth that characterizes a rational person.

If such an account is plausible, then it seems to me that the (act) utilitarian can say the same for Jack as a moral agent. Adams complains that opening a sphere of "blameless wrongdoing" threatens to trivialize the notion of moral wrongness and the profound concern that a good moral agent will show for the rightness of her actions. However, this is only true if we interpret such a concern as an obsession: a concern even for the *tiniest* departures from rightness in the case of the most *trivial* acts.[11] Instead, we might interpret it in a different way. As a good moral agent, Jack will of course be concerned with the rightness of his actions. If he is a utilitarian, this will involve a concern for the way in which his actions contribute to long-run happiness. But he will also be sensible. Just as he understands, as a rational person, that the departure from truth of some of his beliefs is not really worth his concern, so he understands that the minor departure that some of his acts make from optimal utility production does not much matter. This is *especially* true when these acts are done from the best motives he can have.

Despite the efforts of Feldman and others, the fact then remains: Sometimes the best utilitarian motives will lead an agent *not* to perform the best utilitarian acts. This fact, however, while perhaps a source of irony, is not the source of any theoretical conflict, nor is the sphere of "blameless wrongdoing" that it gives rise to obviously implausible.

References

Adams, R. (1976). Motive Utilitarianism. *The Journal of Philosophy, 73*(14), 467–481.

Anomaly, J. (2008). Internal Reasons and the Ought-Implies-Can Principle. *The Philosophical Forum, 39*(4), 469–483.

Bloomfield, P. (2001). *Moral Reality*. Oxford: Oxford University Press.

Crisp, R. (1992). Utilitarianism and the Life of Virtue. *The Philosophical Quarterly, 42*(167), 139–160.

Dancy, J. (1997). Parfit and Indirectly Self-defeating Theories. In J. Dancy (Ed.), *Reading Parfit*. Oxford: Blackwell.

Feldman, F. (1993). On the Consistency of Act- and Motive-Utilitarianism: A Reply to Robert Adams. *Philosophical Studies, 70*(2), 201–212.

Gruzalski, B. (1986). Parfit's Impact on Utilitarianism. *Ethics, 96*(4), 760–783.

Louise, J. (2006). Right Motive, Wrong Action: Direct Consequentialism and Evaluative Conflict. *Ethical Theory and Moral Practice, 9*, 65–85.

Mason, E. (2002). Against Blameless Wrongdoing. *Ethical Theory and Moral Practice, 5*, 287–303.

Mill, J. S. (1979). *Utilitarianism*. G. Sher (Ed.). Indianapolis: Hackett Publishing Company.

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

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

Pettit, P., & Smith, M. (2000). Global Consequentialism. In B. Hooker, E. Mason, & D. Miller (Eds.), *Morality, Rules, and Consequences: A Critical Reader* (pp. 121–133). Edinburgh: Edinburgh University Press.

[11] We might consider here a further analogy. Presumably a good citizen will hold a general and profound concern for the law and the legality of her actions (provided, perhaps, that her society's laws are substantially just). This need not imply a profound concern for the *slightest* violation of even the most *trivial* laws. Even good citizens, it seems to me, will on occasion break the law and not care.

Railton, P. (1984). Alienation, Consequentialism, and the Demands of Morality. *Philosophy and Public Affairs, 13*(2), 134–171.
Streumer, B. (2003). Can Consequentialism Cover Everything? *Utilitas, 15*(2), 237–247.
Streumer, B. (2007). Reasons and Impossibility. *Philosophical Studies, 136*(3), 351–384.
Vihvelin, K. (2015). Arguments for Incompatibilism. Stanford Encyclopedia of Philosophy. E. N. Zalta (Ed.), http://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=incompatibilism-arguments.