

ALASTAIR NORCROSS

SHOULD UTILITARIANISM ACCOMMODATE
MORAL DILEMMAS?¹

(Received in revised form 14 February 1994)

According to some philosophers, a moral agent may find herself confronted with a choice between two or more impermissible or obligatory alternatives. Any choice between impermissible options can be redescribed as a choice between obligatory options. Let us call such alleged situations 'moral dilemmas'. Thus, it might be argued, Agamemnon's tragic choice at Aulis presented him with two options, both of which were impermissible: fail in his paternal duty by killing his daughter; fail in his kingly duty by refusing to appease Artemis. Moral dilemmas have been thought to pose serious problems for ethical theories, in particular for consequentialist theories such as utilitarianism. According to act consequentialist ethical theories, every situation of choice has at least one permissible option. An option is permissible, on standard maximizing theories, just in case it is at least as good as every other option.² In every choice there is at least one option that satisfies this requirement. If utilitarianism is to accommodate moral dilemmas, then, the permissibility requirement must be strengthened. In this paper I will examine two alternative additions to the classical permissibility requirement, the first suggested by Michael Slote, the second by me. In each case, I will argue, the alteration yields far too many dilemmas, unless it gives moral weight to the distinction between making something happen and allowing it to happen. Utilitarians have traditionally denied that this distinction carries any moral significance. I will offer some suggestions as to why this is so, which will amount to a partial defense of the traditional view. I will argue, therefore, that each attempt to accommodate moral dilemmas within a utilitarian framework carries with it too high a price for a utilitarian to pay.

Philosophical Studies 79: 59–83, 1995.

© 1995 Kluwer Academic Publishers. Printed in the Netherlands.

1. MORAL DILEMMAS

Why should utilitarianism make room for moral dilemmas? The most obvious answer is, simply, that they exist. Consider an oft discussed candidate for such a situation: Agamemnon's choice at Aulis. Agamemnon is commander of the Greek forces becalmed on the island of Aulis. The only way to get to Troy is to sacrifice his own daughter. If he doesn't sacrifice his daughter, all the Greeks are stuck, and his brother Menelaus is denied the opportunity to get his wife back and exact vengeance on the Trojans for Paris's crime. Well, obviously Agamemnon, as commander of the forces and brother of the cuckold, must do what he can to get from Aulis to Troy. It *can't* be right for him not to do what is required to accomplish this. But, obviously it *can't* be right for Agamemnon to sacrifice his own daughter, Iphigeneia. So, whatever he does or does not do, however he behaves, he will have acted, or behaved, wrongly. The alternative is to maintain that his situation *is* tragic, that he has to choose between two highly unpleasant alternatives, each of which will be accompanied by guilt, regret and remorse, but one of which is right and the other wrong (ignoring the highly unlikely event that they were exactly equal). Let us see if there are any reasons to adopt one of these alternatives rather than the other.

Does 'Agamemnon acts wrongly in doing x' entail 'Agamemnon ought not to do x'? Let us first suppose that it does. In this discussion, doing x is sacrificing Iphigeneia, and doing y is allowing the Greek force to remain stranded on Aulis. Further, it is stipulated that (i) the only alternative to doing x is to do y, and thus that (ii) if Agamemnon does not do x then he does y. If this situation is dilemmatic, then Agamemnon acts wrongly in doing x and he would act wrongly in doing y; so he ought not to do x and ought not to do y. But since his not doing x entails, with (ii), his doing y, it follows that he ought not to do y and ought to do y.³

This is not a formal contradiction, but it is still hard to accept. How can it be true both that I ought to do y and that I ought not to do y? The proponent of moral dilemmas might simply reply that it can be true, that Agamemnon's situation illustrates this and that that is what a moral dilemma is. I am reluctant to be satisfied with such

a response, without first attempting to explain how situations such as Agamemnon's can be tragic, but not such that he ought to do *y* and ought not to do *y*. The claim that Agamemnon both ought to have allowed the Greeks to remain stranded and ought not to have done so is, as I said, highly counterintuitive. This fact should at least warrant a search for an alternative explanation for our intuition that his choice was tragic. Such a search is not arduous. Traditional versions of utilitarianism can make perfectly good sense of the idea that Agamemnon's choice is tragic, without maintaining that he ought to have allowed the Greeks to remain stranded and he ought not to have done so. The choice is tragic because each alternative is highly unpleasant (to put it mildly), each alternative leads to disastrous consequences. But is there no difference between the alternatives?

In discussing situations like this, I have occasionally encountered in others an attitude which I find very puzzling. When a choice is described as being between two alternatives, each of which has disastrous consequences, it is sometimes suggested that it is impossible to choose between the two on the basis of any principle, because each is simply too horrible to contemplate. Thus, it might be suggested that the choice between killing a million and killing one and a half million people is not a moral choice. This puzzles me because it seems clear that if it is at least in principle possible to compare two situations with respect to how bad they are, then it is at least in principle possible to make a reasoned choice between them. It seems reasonable to suppose that one of Agamemnon's choices was worse than the other. This supposition in no way diminishes the awfulness of the situation in which Agamemnon found himself, and is compatible with our normal reactions to the story. Of course, even if both choices were exactly equal, they would not both be wrong, on a traditional utilitarian account, unless there were a third choice, better than both of them.

On the other hand, let us suppose instead that 'Agamemnon acts wrongly in doing *x*' does not entail 'Agamemnon ought not to do *x*'. This allows for the possibility that Agamemnon acts wrongly in doing *x* and in doing *y*, and yet he ought to do *x* (say) and not to do *y* (and it is not the case that he ought not to do *x*). This is counterintuitive too. How can it be that someone morally ought to do what is morally wrong for

them to do? To divorce what it is wrong for an agent to do from what she ought not to do seems to involve changing the meaning of 'wrong'.

2. SLOTE ON DILEMMAS

The concerns raised in the last section may motivate us, especially those of us who are utilitarians, to deny the existence of moral dilemmas. The intuition that some situations are truly dilemmatic may, however, persist, even for a utilitarian. Michael Slote has recently suggested a version of act-utilitarianism which allows for the existence of moral dilemmas. The traditional utilitarian criterion of a morally right action is that it produces as much good overall as any of the alternatives available to the agent in a given situation. If there are two actions which tie for first place, they are both morally right. Slote argues that it is less plausible to say this of actions "with overall bad consequences that are nonetheless better than anything else an agent can do in given circumstances"⁴ than of actions with overall good consequences. He doesn't attempt to justify this, but uses it to construct a version of act-utilitarianism which does not judge any action morally right which does not make "some contribution on balance to human happiness".⁵ Slote interprets the original edition of Bentham's *Introduction* as understanding "the principle of utility as making the rightness or wrongness of an act depend on whether it contributes to human happiness or to human unhappiness."⁶ He combines this demand that any right act have overall good consequences with the further demand that any right act maximize good consequences, and produces "the moral theory that results if one demands of a right action both that it produce consequences no less good than those producible by any alternative act available to a given agent and that those consequences be, on balance, good."⁷ This gives the following account of rightness:

- UD An act is right iff (i) it produces consequences no less good than those producible by any alternative act available to the agent; and (ii) it produces consequences that are, on balance, good.

This theory, says Slote, entails the “possibility of situations where no matter what one does, one acts wrongly.”⁸ That is, it entails the possibility of moral dilemmas.

An example of the sort of case Slote has in mind might be Williams’s case of Jim and the villagers. If Jim doesn’t kill one of twenty villagers, someone else will kill all twenty. This is the kind of case that strikes many⁹ as a paradigmatic moral dilemma. No matter what he does, Jim does something wrong. Part (or even all) of the motivation for this judgment might be the belief that someone must be wronged in all of Jim’s best alternatives. One might reason as follows: Since Jim can easily prevent nineteen deaths, it would be wrong for him not to do so. But the only methods available to Jim to prevent the deaths involve wronging a twentieth. So Jim does wrong no matter what he does. This line of thinking clearly conflates two different senses of ‘wrong’. It may be possible to wrong someone without doing wrong, and to do wrong without wronging anyone. It is, nonetheless, a powerful source of the intuition that Jim is faced with a moral dilemma.

According to Slote, his version of utilitarianism has as good a claim “to express the fundamental moral ideals of utilitarianism as contemporary versions.”¹⁰ Utilitarianism takes a universalized form of ordinary human benevolence as “the touchstone of moral justification. . . . Ordinary benevolence can express itself no less in the desire to do well by people than in the desire to do the best one can for people”.¹¹ There may be situations in which anything one does or does not do will make things worse for people, in which case the desire to do the best one can for people may be satisfied, but the desire to do well by people may be thwarted. Why, asks Slote, should utilitarianism require the satisfaction of only one component of universal benevolence?

At this point it might be objected that Slote’s proposal entails that intuitively indifferent actions will be deemed impermissible. The demand that an action produce good consequences rules out any actions that simply don’t affect the welfare of morally relevant beings. Slote could reply that this isn’t any more of a problem for his modification of utilitarianism than for the traditional version of the theory. After all, he might say, the demand that an action be the best of the available alternatives also rules out actions that are intuitively indifferent, when

there are better alternatives available to the agent. So the utilitarian, to whom Slote addresses his proposal, shouldn't be particularly concerned by this criticism. However, it is worth noting that Slote's proposal rules out *all* intuitively indifferent actions, while traditional maximizing utilitarianism only rules out some such actions. If an intuitively indifferent action is one of an agent's best alternatives, it is permissible, and if it is the unique best, it is required, according to traditional utilitarianism. There is, of course, a simple modification to Slote's thesis that would solve this special problem (as opposed to the general problem shared with traditional utilitarianism that some intuitively indifferent actions are judged impermissible). If, instead of demanding that an action produce overall good consequences, he demands that an action not produce consequences that are, on balance, bad, some intuitively indifferent actions may be permissible. In fact, at some points of his paper Slote seems to interpret his suggestion in precisely this less demanding way. It is, therefore, to this suggestion that I turn in the next section.

3. MAKING THINGS WORSE

Slote says that ordinary benevolence expresses itself in the desire to 'do well by people'. Later, he claims that this desire is the desire "of actions . . . that they not make things worse than they already, imperfectly, are."¹² If we modify the second requirement of UD to fit this suggestion, we get:

UD' An act is right iff (i) it produces consequences no less good than those producible by any alternative act available to the agent; and (ii) it produces consequences that are not, on balance, bad.

As I made clear in the previous section, the demand that an action not make things worse is weaker than the demand that it have overall good consequences. It still seems to be a significant addition to traditional utilitarianism, however, so let's look at the theory that results from this addition. The first thing to note is that it is questionable whether doing well by people requires not making things worse than they already are.

If I am faced with a control panel covered with buttons which regulate the amount of pain to be suffered by a victim, and am told that if I press nothing, the victim will die painfully, do I not do well by him by pressing the button which causes the least pain?

Besides, the idea of an action making things no worse than they already are is problematic. Consider the following example: a lighthouse keeper, called 'Keeper', sees a shipwreck which results in hundreds of people being plunged into the icy sea. He knows that they will die soon if they are not rescued, and so he must act fast. They are not all in the same place. Three-quarters of them are north of the lighthouse, while the other quarter are south. If Keeper goes north and rescues the larger group, the smaller group will be dead by the time he can get to them. If he goes south first, the larger group will be dead by the time he can get to them. No-one else is available to help. Keeper goes north and rescues the larger group. The smaller group all die. Recall that we are considering a version of utilitarianism that demands of a right action both "that it produce consequences no less good than those producible by any alternative act available to a given agent" and that it "not make things worse than they already are". Keeper's action satisfies the former constraint, but what about the latter?

What is it for an action to make things worse than they already are? Here's a possibility: an action makes things worse than they already are iff the consequences of the action are, on balance, bad. The demand that an action not make things worse, on this interpretation, differs from the demand that an action have consequences that are, on balance, good only in that it allows an action to have consequences that are, on balance, neither good nor bad. Are the consequences of Keeper's action, on balance, bad? The consequences in which a utilitarian is interested concern the welfare levels of all those affected by the action. Those affected by the action are the survivors of the shipwreck. Both groups are affected by Keeper's action. Most of the people are better off after he takes his boat out, since they were wet, cold and frightened before, but warm, dry and tranquil after. Some, however, are dead after. The drop in welfare for the few is much bigger than the rise for the many. If we try to compare the situation of all those affected by Keeper's action before he took his boat out with that after he completed his rescue, the

later situation is worse than the earlier one, since more people are dead later.

If my example accurately illustrates Slote's proposal, it would appear that moral dilemmas are somewhat in abundance. There are plenty of cases in which the overall level of welfare of those affected by my action drops, even though my action is at least as good as any available alternative. There are, I'm sure, many people right now, each of whose welfare level is declining in such a way that I could do something to slow, but not halt or reverse the decline. In fact, I'm inclined to think that the situation is far worse with respect to the proliferation of dilemmas according to this reading of Slote's proposal. There are literally millions of people right now, each of whose welfare level is declining in such a way that I could do something to slow the decline, and, for some I *could* halt or even reverse the decline. Most of these people are in distant countries, but a significant number are in this country. Each of my best alternatives, however, involves helping a small minority of all of these people. So, in each of my best alternatives the overall level of welfare of those affected by my action declines. This is so, because no matter how many people I help, there are far more people, each (but not all) of whom I could have helped instead. The decline in the level of welfare of the latter people vastly outweighs the increase, if any, in the level of welfare of those I manage to help. This is to say that I (and perhaps most people in the developed countries) am in the same situation as Keeper with respect to the most significant choices I face. Even a friend of dilemmas, let alone a utilitarian, should be uncomfortable with the result that almost every significant choice I make involves a moral dilemma.

Perhaps my example doesn't really illustrate Slote's proposal at work. It might be objected that Keeper doesn't *make* things worse than they already are. He merely *allows* things to become worse. The same will be true of most cases in which my best option involves a drop in the welfare of those affected. When I help some people, and thereby fail to help others, I don't make the latter poorer, sicker, more miserable or more dead than they were before. I allow them to become so. Keeper doesn't actually kill the people in the smaller group. He lets them die. If we have an account of moral dilemmas which is sensitive to the distinc-

tion between making something happen and allowing it to happen, we may be able to avoid the result that almost everything we do involves us in moral dilemmas. I will have more to say later about the distinction between making and allowing. For now, let me point out that utilitarians usually deny that it has any moral significance. If we incorporate the distinction between making and allowing into our version of utilitarianism, we will demand, not that the consequences of every action be not, on balance, bad, but only that every action not positively make things worse. This gives us:

UD'' An act is right iff (i) it produces consequences no less good than those producible by any alternative act available to the agent; and (ii) it does not positively make things worse.

The second requirement of UD' is stronger than the second requirement of UD'' in that the former demands that every action neither make things worse nor allow them to become worse.

In my example, Keeper is faced with a situation in which, whatever he does, things will become worse. However, his best option does not involve making things worse, only allowing them to become worse. On a version of Slote's suggestion that is sensitive to the making/allowing distinction, Keeper is not faced with a moral dilemma. But it is possible to modify my example so that it fits the altered description of a moral dilemma. This time the survivors of the shipwreck are all in the same area. Keeper can get to them all before they drown or freeze to death. It is dark, however, and he knows that if he takes his lifeboat out to rescue the survivors, he will kill some of them. Obviously Keeper takes his boat out, since if he didn't, they would all die. As he predicted, he kills some of the survivors in the water, but he rescues most of them. Once again, most of the people are better off after he takes his boat out, but some are much worse off. The overall level of welfare of those affected by his decision drops. This time, however, Keeper makes the level go down. He kills the people whose drop in welfare accounts for the overall drop.

If my second Keeper example counts as a moral dilemma according to Slote's account, we should reject the account. Remember that Slote is addressing utilitarians who are toying with the idea of accommodating

dilemmas. I don't deny that some nonutilitarians who are already sympathetic to dilemmas will judge that Keeper is faced with a dilemma. After all, they might say, he has to kill innocent people to save more innocent people. Isn't that, like the case of Jim and the villagers, a paradigmatic dilemma? In fact, even among the friends of dilemmas, there isn't a clear consensus on this point. There are those who think it significant that the one villager would also die if Jim refused to shoot. (This may be why Williams himself admits that the utilitarian decision for Jim to shoot the one villager is the correct one.) Similarly, they might say, those who are killed by Keeper would also die if he didn't take his boat out. On the other hand, they might continue, if Keeper (or Jim) had to kill some people *who wouldn't otherwise die*, in order to save other people, that would clearly be a moral dilemma. So if Keeper is faced with a moral dilemma, according to Slote's proposal, even many of those who are already sympathetic to dilemmas will have reason to reject the proposal. More importantly, a utilitarian who is considering modifying her theory would not want a modification that would classify Keeper's situation as a moral dilemma.

An even more dramatic illustration of why a utilitarian should reject the latest version of Slote's proposal is provided by the following case: a doctor, named "Doctor", is caring for a patient named "Patient". Patient is terminally ill. His condition is declining, and his suffering is increasing. Doctor cannot delay Patient's death. The only thing she can do is to slow the rate of increase of Patient's suffering by administering various drugs. The best available drugs completely remove the pain that Patient would have suffered as a result of his illness. However, they also produce, as a side-effect, a level of suffering that is dramatically lower than he would have experienced without them, but significantly higher than he is now experiencing.¹³ So the result of administering the drugs is that Patient's suffering continues to increase, but at a slower rate than he would have experienced without them. Doctor's situation fits Slote's suggestion. The account of moral dilemmas we are considering is based on two requirements: (i) An agent must do the best she can; (ii) An agent must not positively make things worse. Doctor can satisfy (i), but only at the expense of (ii). Her best option is to administer drugs to Patient. In this case, though, Patient's condition is the direct causal result

of Doctor's action. Once again, many proponents of moral dilemmas would judge this case to involve a dilemma. The only way that Doctor can reduce Patient's suffering is to make him suffer herself. The main purpose of the case, however, is to pressure the friend of utilitarianism who is toying with modifications to the theory to allow for dilemmas. A utilitarian would clearly not think that Doctor is faced with a dilemma. Doctor is minimizing Patient's suffering.¹⁴

It might be objected that my rejection of the latest version of Slote's proposal relies on a mistaken account of what it is to make things worse than they already are. At first blush it doesn't appear that either Keeper or Doctor really do make things worse than they already are, at least not in any sense that would appeal to a utilitarian. Perhaps we can say that they make things better, since before he took his boat out, all the people were *going to die*, and before she gave Patient the drugs, he was going to suffer even more. But the people were only going to die if Keeper didn't take his boat out, and Patient was only going to suffer even more if Doctor didn't administer the drugs. Perhaps we should say that Keeper and Doctor make things worse than they already are, but not worse than they were going to later be. After all, we might think, we sometimes make things worse than they now are in order to make them better than they would otherwise wind up being. As it stands this suggestion is ambiguous between at least the following two possibilities: (i) As a result of a particular action, the welfare level of those affected drops, in the short term, below what it would otherwise have been, but in the long term the level is higher than it would otherwise have been. (ii) As a result of a particular action, the welfare level of those affected drops, but it remains at all times at least as high as it would otherwise have been, and at some times higher.

(i) seems to fit my second Keeper example. The short term drop in welfare occasioned by Keeper's rescue is greater than if he hadn't gone out. He kills some people who wouldn't have been dead in the short term. But those people and all the others would have been dead in the slightly longer term. So the long term drop is smaller than if he hadn't gone out. (Of course, if the few who are killed by Keeper were actually better off dead than they would have been for the few minutes before they drowned, this would be a case of (ii).) (ii), on the other hand, fits

my Doctor example. Patient's welfare level drops as a result of Doctor's action, but it is, until his death, higher than it would otherwise have been. (After his death it is, of course, the same as it would otherwise have been.) A utilitarian should not object to the claim that actions of type (i) make things worse (at least in the short term). But it is not at all clear that a utilitarian should say that actions of type (ii) make things worse in any sense. The temptation to say that Doctor makes things worse comes from comparing the welfare levels of Patient at different times in the world in which Doctor administers treatment. But cross-temporal comparisons are not the relevant ones for utilitarian evaluations. The utilitarian is interested in comparisons across worlds, not across times. In the next section I will explore the possibility of a utilitarian account of making things worse that uses cross-world comparisons.

4. GOOD AND BAD CONSEQUENCES

In order to decide whether something makes things worse than they already are, we can't simply compare states of affairs at one time and at a later time. We compare the results of the different choices of behavior with each other, not with the situation before the behavior. There is no need to compare the world which results from doing x with the current state of the world, and then perform a similar comparison for y. When a utilitarian is trying to decide whether to do x or y, she compares the consequences of doing x with those of doing y. An action with overall bad consequences makes things worse, not than they were, but than they *would have been* if the action hadn't been performed.

The current proposal is that a utilitarian judges whether an action has overall good or bad consequences (or neither) by comparing the world that results from it with the world that would have resulted if it hadn't been performed. The latter world provides a neutral point in terms of the value of the consequences. (Since I am considering this proposal about the value of an action's consequences as part of Slote's proposed modification of utilitarianism, I should stress that an action with neither good nor bad overall consequences would not necessarily be morally neutral on that account. There is still the maximizing condition to be

met.) Any action that results in a better world has good consequences. Any action that results in a worse world has bad consequences. There are problems with the proposal, though. When we wonder what the world would have been like if a particular action hadn't been performed at time *t*, how do we imagine things to have been changed? Two approaches with *prima facie* plausibility can be easily dismissed: (i) We imagine the agent to have remained immobile at *t*; (ii) We imagine the agent to have been absent from the scene at the *t*. Within (ii) there are two possibilities: (ii)a We imagine a world identical to the actual world before *t*, in which the agent miraculously vanishes from the scene at *t*; (ii)b We imagine a world as similar as possible to the actual world before *t*, in which the agent is non-miraculously absent from the scene at *t*. That is, we imagine what would have had to have been different before *t* in order for the agent to have been absent at *t*.¹⁵ Both immobility and absence from the scene fail to give us a satisfactory neutral point with which to compare the results of an agent's actions. Not only do they give the wrong moral judgments about some actions, but, perhaps more importantly, they don't seem to capture what we mean when we ask what would have happened if the action hadn't been performed. The comparisons they invite are simply not the correct ones. When we wonder what would have happened if we hadn't done something, we don't suppose ourselves to be immobilized, or to be removed from the scene, miraculously or not.

If we are to judge an action by comparing its consequences with what would have happened otherwise, we need an intuitively acceptable account of the latter notion. The best account I know is to be found in the writing of Alan Donagan. Donagan defines an action as "a deed done in a particular situation or set of circumstances; . . . [consisting] partly of [the agent's] own bodily and mental states".¹⁶ He continues:

Should he be deprived of all power of action, the situation, including his bodily and mental states, would change according to the laws of nature. His deeds as an agent are either interventions in that natural process or abstentions from intervention. When he intervenes, he can be described as causing whatever would not have occurred had he abstained; and when he abstains, as allowing to happen whatever would not have happened had he intervened. Hence, from the point of view of action, the situation is conceived as passive, and the agent, qua agent, as external to it. He is like a *deus ex*

machina whose interventions make a difference to what otherwise would naturally come about without them.¹⁷

In considering what would have happened if an agent hadn't acted, Donagan doesn't imagine her to be immobile or absent from the scene. Instead, he asks what would have happened in "the course of nature" (his phrase). The course of nature can include not only the agent's physical presence, but also changes in her "bodily and mental states". It is the exercise of human agency that gives an agent the option to intervene in the course of nature or to allow nature to take its course. All of an agent's deeds are either interventions or abstentions. Those that make a difference to the course of nature, or what would have happened anyway, are interventions; those that leave the course of nature unchanged are abstentions. Whatever problems there are with this analysis of the difference between doing and allowing,¹⁸ it seems to have a good deal of intuitive support. How, then, does this account fare as part of an explanation of what it is for an action to have good or bad consequences?

The major problem with this latest suggestion is that it entails that letting nature take its course never has either good or bad consequences overall. If someone allows to happen what would have happened anyway, the consequences of her behavior are morally neutral (though not necessarily the behavior itself). This may be more amenable to anti-consequentialists, who are more inclined to invest the doing/allowing distinction with moral significance than are consequentialists, but it will be unacceptable nonetheless. Even those who claim that it is generally worse to do bad things than to allow them to happen (and that it is better to do good things than to allow them to happen?) will admit that some allowings can have very bad consequences and some can have very good ones. One example will suffice: Beth is a student who stumbles onto an experiment conducted by a twisted scientist, named Scientist. He is seated at a control panel containing several buttons, beyond which are ten people strapped to chairs and connected to electrodes. In a few moments, he explains, a red light will flash, after which, if no button is pressed, all ten of his victims will die slow lingering deaths. Each button, if pressed, will cause a different section of his victims to be tortured to a different degree. He was about to press the button that would have

caused slow lingering deaths to all ten. However, to honor her arrival, he turns control of the buttons over to Beth. She doesn't have to push the worst button. The second-worst, for example, only kills eight people, the third-worst six, and so on. She need only press one button to prevent the wholesale slaughter of the victims, and the least harmful button will not even kill anyone, but merely torture a couple of healthy young football players for a few minutes at a fairly low intensity. It is hardly plausible to claim that the consequences of pushing no buttons are morally neutral. Worse still, according to this account, the consequences of pushing the second-worst button would be overall good.

I began this paper with a discussion of some general considerations concerning whether to accommodate moral dilemmas within a moral theory. I then turned to Slote's suggestion that we add a second requirement to the traditional maximization requirement of utilitarianism. The first version of Slote's suggested addition – that an action have consequences that are, on balance, good encountered the objection that all intuitively indifferent actions were rendered impermissible. A slightly weaker version – that an action have consequences that are not, on balance, bad – gave far too many dilemmas even for a friend of dilemmas, let alone a utilitarian, to countenance. A reading of Slote's condition as sensitive to the difference between doing and allowing solved the problem of proliferation of dilemmas. This reading, however, besides the obvious problem of incorporating a distinction inimical to utilitarians, rendered as dilemmatic the sort of choices they should find unproblematic. In the current section, I have been criticizing the suggestion that a utilitarian evaluate the goodness or badness of an action's consequences by comparing the world that results from the action with the world that would have resulted if the action hadn't been performed. In the next section I will consider the suggestion that we replace Slote's requirement not to make things worse with the requirement not to do ill by anyone.

5. A DIFFERENT APPROACH¹⁹

Slote's alteration of the permissibility requirement springs from the suggestion that what underlies utilitarianism is really the desire to do

well by people. This desire, he claims, generates the requirement that our actions not make things worse. However, the desire to do well by people might more naturally generate a requirement not to do ill by someone. This line of thought might render the following account of a moral dilemma: it is a situation in which there is nothing an agent can do that satisfies both the requirement to do the best she can and the requirement to avoid ill-doing.

What is it to do ill by someone? Let's start with the idea of someone undergoing something very bad. This requires that we have not only judgments of the form "x is better than y" but also judgments of the form "x is very bad indeed". We often do make judgments of that form about the suffering of others (or ourselves). So the idea of doing ill by someone is

So conducting oneself that the other undergoes something very bad indeed, when one could have behaved in such a way that the other didn't undergo anything as bad.

This suggestion suffers from the same problem as my initial reading of Slote's suggestion. It seems to produce far too many moral dilemmas. In fact, most of us, at least in the West, are most of the time conducting ourselves so that there is at least one person who undergoes something very bad indeed, who would not undergo anything as bad had we behaved in a certain different way. No matter what I'm doing, even if I'm saving the lives of children and kittens caught in a burning building, there's probably someone somewhere whose suffering I could be alleviating instead. Even worse, on this proposal I would do you ill if I fail to reduce your very bad suffering by a tiny amount, even if the reduction could only be accomplished at the cost of enormous suffering to myself or others.

Perhaps we could again avoid the proliferation of dilemmas by incorporating the making/allowing distinction into the concept of ill-doing. We need a concept that is asymmetrical with respect to the difference between making someone suffer and allowing them to suffer. Consider the following asymmetrical concept:

So conducting oneself that one causes the other to undergo something very bad indeed, when one could have behaved in such a way that the other didn't undergo anything as bad.

My use of “cause” here is intended to distinguish what one causes from what one allows to happen, even though the latter is among the causal upshots of one’s behavior. If we employ something like this asymmetrical concept of ill-doing in our characterization of moral dilemmas, we avoid the problem of proliferation. Consider the following notion of a moral dilemma:

DA A moral dilemma is a situation in which anything the agent can do is either an instance of ill-doing or has worse consequences than some other action available to the agent.

The first disjunct of this proposal involves a concept that is sensitive to the distinction between making and allowing. The second disjunct reflects the requirement to optimize, which is not sensitive to the making/allowing distinction. The consequences of our behavior include all the things we allow as well as the things we make happen. On this account, if all of the best options open to an agent involve doing ill by someone, then the agent will be in a moral dilemma. A variation on Bernard Williams’s case of Jim and the Villagers will serve as an example.²⁰ If Jim must either shoot this Villager or allow Pedro to shoot the other nineteen, but not this one, he is in a dilemma.²¹ The only alternative to doing ill by the Villager is allowing worse things to happen.

Perhaps DA will produce moral dilemmas in roughly those situations where they are commonly supposed to exist. If so, it is significant that we have been able to get this result from within a utilitarian framework. Should a utilitarian, or any other moral theorist, adopt DA? The question again arises as to the concept of wrongness employed here. If it is wrong either to do ill by someone or to fail to optimize, there will be situations where an agent can’t help acting wrongly. If our moral theory demands that we don’t act wrongly, then it issues commands which do, on occasion, conflict. That is, sometimes the demands of morality *cannot* be met; it is not just that they are hard to meet. Why should we accept such a morality? If, on the other hand, our notion of what is morally wrong does not coincide with our notion of what morality

demands that we don't do, then it is not clear that we have a recognizable concept of wrongness.

The concerns expressed in the previous paragraph apply to any attempt to argue for the existence of moral dilemmas. There are further questions that arise about DA. DA employs a concept of ill-doing that is asymmetrical with respect to the difference between making someone suffer and allowing someone to suffer. Moral dilemmas only occur in situations where an agent's best option involves ill-doing. Consider two situations in each of which one course of action has much better consequences than any other that the agent could have chosen, this best option resulting in someone's undergoing something very bad indeed. Both situations are variations on one of Judith Thomson's examples. Let us call the first situation "Make". In Make there is a train thundering out of control down a track towards five innocent people who will be killed if the train isn't stopped. A small distance in front of the five innocents is a bridge across the track, on which is precariously perched a very large man, Mega. He is deep in thought and easily startled. If someone were to shout "Mega" at the top of her lungs from nearby, Mega would fall from the bridge into the path of the oncoming train and bring it to a halt. He would be killed by the combination of the fall and being struck by the train. Nothing but Mega's body can stop the train before it kills the five innocents. Nearby is a small but voluble woman, Pusa. She is aware of all the relevant facts concerning the situation. Her best option is to shout "Mega". This will involve doing ill by Mega. All her other options are worse, since they involve the deaths of the five innocents. Let us call the second situation "Allow". Allow is very similar to Make. There are a couple of significant differences. Mega is still perched precariously and deep in thought. He would not, however, fall off the bridge if Pusa were to shout "Mega". He would look up, notice the swooping bird that is about to knock him off the bridge, and avoid falling into the path of the oncoming train. So now Mega will fall into the path of the train and prevent it from killing the five innocents, *unless* Pusa shouts "Mega". Pusa is aware of all this. Her best option is not to shout "Mega", but to allow him to be knocked off the bridge by the bird. Let us say that Pusa selects her best option in both Make and Allow. In each case Pusa's action has the bad result that Mega dies (and the good result that the five

innocents live). The situations differ only in that in one of them this bad result is made to happen, while in the other it is allowed. According to DA, Make is a moral dilemma, but Allow is not. If DA is to be of any moral interest, there must be a morally significant difference between Make and Allow. We would not, I presume, say that although Make is a moral dilemma and Allow is not, they are equally morally bad, or unpleasant, or distressing. The difference between making and allowing must therefore be morally significant. Furthermore, it would seem that it must be, at least sometimes, morally worse to make something bad happen than to allow it, *even if all other things are equal*.

DA attaches moral significance to the making/allowing distinction, but only in the requirement not to do ill. The requirement to optimize is neutral with respect to the distinction. This seems inconsistent. If there is a morally significant distinction between making and allowing, shouldn't this be reflected in all the requirements of morality? Perhaps optimization is special. It might be said that it usually makes a difference to the moral value of my action whether I make or merely allow certain consequences to obtain, but not if those consequences are at least as good as those of any available alternative. There is something so special about the best possible consequences that it doesn't matter whether I positively make them occur or merely allow them. I don't think that this is plausible. There seems to be no more reason to say that the best alternative should be exempt from the making/allow distinction than that the top three should be, or all but the worst.

If we want to avoid inconsistency, perhaps we can modify the requirement to optimize so as to incorporate the distinction. What is it to do the best I can? This is the usual understanding:

OP I do the best I can iff the consequences of my action are no worse than those of any available alternative.

If there is a morally significant distinction between making and allowing, it is worse to make something bad happen than to allow it, and it is better to make something good happen than to allow it. So the moral value of my action is determined not solely by the value of its consequences compared with the consequences of the alternatives, but by the value of its consequences combined with certain facts concerning my

relation to those consequences. We need to adjust the value of the consequences of my action depending on whether I make those consequences obtain, or merely allow them to obtain. This will give us the notion of the *adjusted value* of consequences.²² We can use this notion to give an optimizing requirement that incorporates the making/allowing distinction:

OP' I do the best I can iff the adjusted value of the consequences of my action is no lower than that of any available alternative.

Exactly what adjustments we make to the value of the consequences will depend on what significance we accord to the making/allowing distinction. For example, we might say that it is twice as bad to make something bad happen as to allow it to happen, and twice as good to make something good happen. The significance might be more complicated than that, though. Perhaps the difference increases as the consequences get worse, or certain kinds of bad consequences are far worse to make happen than to allow.

6. UTILITARIANISM AND BENEVOLENCE

The questions of whether and to what extent there is a morally significant difference between making and allowing are beyond the scope of this paper.²³ I will say here that it is at least a controversial issue. As I said above, utilitarians are inclined to reject the moral significance of the distinction, even if they accept its metaphysical reality. It may be instructive to see why this is so. A consequentialist theory such as utilitarianism consists of a value theory – a theory of what is good or bad – and a function that tells us what actions are right and wrong. The most common function from the good to the right can be expressed as follows: an act is right iff it results in at least as much good as any alternative available to the agent, otherwise it is wrong. According to the classical utilitarian value theory, happiness is the only thing that is intrinsically good, and unhappiness the only intrinsic bad. A utilitarian is concerned that there be as much good and as little bad as possible. Slote is correct in his claim that utilitarianism takes a universalized form of

ordinary human benevolence as “the touchstone of moral justification”. The essence of ordinary benevolence, though, is neither the desire to do well by people nor the desire to do the best one can for people. It is the desire that people be well off.²⁴ In a world full of very happy people a benevolent person is not thwarted by the inability to make people happy. She is delighted that her efforts in that direction aren’t needed. A benevolent person desires to do well by people only because she desires people to be well off. In the actual world, for the most part, the best way to satisfy the latter desire is to satisfy the former. It is important to see that the desire to do anything is only a contingent feature of benevolence. It is only because there is so much happiness to be produced that benevolence expresses itself in the desire to do well by people.

It might be objected that the benevolent person gains satisfaction from benefiting people, and that the disposition to gain such satisfaction is a necessary feature of benevolence.²⁵ Furthermore, the satisfaction gained from benefiting people results from the satisfaction of the desire to benefit people. But I am not claiming that benevolent people do not desire to benefit others, only that such a desire is not an essential feature of benevolence. Consider Benvolio, a truly benevolent person. He lives in a clearing in a thick jungle. One day he sees a badly injured airman parachuting towards his clearing. Benvolio has a fair amount of medical training, and knows that he can patch up the airman fairly well. He probably won’t be able to set the broken bones perfectly or treat the burns optimally, but the airman will live, albeit scarred and limping. Benvolio hopes fervently that the airman doesn’t get blown into a different part of the jungle, where he will be unable to rescue him. He wants to do well by the airman.

What if the situation were slightly different, though? Suppose that there is a two doctor team, comprised of a highly trained burns specialist, Skin, and a highly trained orthopedic surgeon, Bones, living in a clearing a few hundred yards away from Benvolio, on the opposite side of an uncrossable river. Doctors Skin and Bones are conducting experiments on the use of advanced surgical techniques in the jungle. They will treat the airman if they can, not because they want him to be well off, but because they want to practice their skills. Benvolio knows that

the airman would receive much better care if he dropped near Skin and Bones. In this case, he fervently hopes that the airman comes down into the doctors' clearing. His benevolence is primarily concerned with the well-being of the airman, not with *how* the airman is benefited. Benvolio's benevolence doesn't require that *he* help the airman,²⁶ if someone else can do a better job, nor even that the airman be helped by anyone, if he could make a better recovery without the exercise of anyone's agency. If there is a third clearing with magical healing properties, where the airman would recover completely without anyone's help, Benvolio will hope most of all that the airman come down there.

The distinction between making someone happy or unhappy and allowing them to be so concerns the method by which good or bad states of affairs are brought about. If I am right about the nature of benevolence, it should not be sensitive to this distinction. To the extent that utilitarianism is based in benevolence, it too will not be sensitive to the distinction. A modification of utilitarianism may, of course, be constructed in which a sensitivity to the making/allowing distinction is built into the function from the good to the right.²⁷ Such a move would require independent justification.

7. CONCLUSION

It is clear that classical utilitarianism does not allow for moral dilemmas. If a right act is one of the best acts available to an agent, it is always possible to avoid acting wrongly.²⁸ I have been exploring the possibility of strengthening the permissibility requirement so as to allow for situations in which no alternative is permissible. Slote suggests that a right act must also have consequences that are, on balance, good. I have attempted to explain what could be meant by this suggestion. One interpretation involves comparing states of the world before and after the action, as opposed to comparing different possible future states. This allows for moral dilemmas, but produces too many, unless a distinction is introduced between making bad things happen and allowing them to happen. Even with the distinction, this suggestion classifies as moral dilemmas many situations which clearly are not. Perhaps more impor-

tantly, the method of comparing past with future states of the world produces a counterintuitive version of making things worse. No interpretation that involves comparing the consequences of an action with those of possible alternatives fares any better. My suggestion involves the notion of doing ill by someone. Once again, this produces too many dilemmas, unless the making/allowing distinction is incorporated. Each suggestion, then, involves giving moral weight to a distinction that is unacceptable to utilitarians. Furthermore, it is far from clear whether ordinary benevolence, in which utilitarianism is based, is sensitive to the distinction. I think it likely that any plausible dilemma-allowing morality will have to give moral weight to the making/allowing distinction, but that is the subject of another paper.²⁹ I can find no other ways, with even *prima facie* appeal to utilitarians, of incorporating moral dilemmas within a utilitarian framework. I conclude that it is at best highly unlikely that utilitarianism can accommodate moral dilemmas without incorporating a distinction that violates the spirit of the theory.

NOTES

¹ I am grateful to Jonathan Bennett and Frances Howard-Snyder for extensive comments on earlier versions of this paper, and to Douglas Ehring and Steven Sverdlik for comments on more recent versions. I am also especially grateful to an anonymous referee for this journal whose many detailed and helpful comments have, I hope, improved the final version considerably.

² Alternatively, an option is permissible just in case there are no better options. These conditions are equivalent, unless options are incomparable. If some options are incomparable, it may be true of a particular act that there are no better options, but false of it that it is at least as good as every other option. I assume in this paper that all options are comparable. The question of value incomparability is the subject of another paper.

³ This conclusion relies on the principle that if S ought to see to p and it is unalterable for S that if she sees to p then she sees to q, then S ought to see to q. This principle is highly plausible, but, as with many principles of deontic logic, there may be those who wish to reject it.

⁴ Michael Slote, "Utilitarianism, Moral Dilemmas, and Moral Cost," *American Philosophical Quarterly*, Vol. 22. No. 2, April 1985, p. 161–68, at p. 162.

⁵ Ibid.

⁶ Ibid.

⁷ Ibid.

⁸ Ibid. pp. 162–3.

⁹ Though, interestingly enough, not Williams himself, who admits that killing the one villager is probably the right thing for Jim to do. Williams's quarrel with the utilitarian concerns the decision procedure Jim is supposedly committed to by utilitarianism.

¹⁰ Ibid. p. 163.

¹¹ Ibid.

¹² Ibid.

¹³ Why not simply have an example involving a regular pain-killer, that removes some, but not all, of the pain that Patient would have suffered? In order for this example to count as a dilemma on the current version of Slote's proposal, Patient's later pain has to be caused by Doctor administering the drugs. According to some intuitively appealing accounts of causation and mental state identity, ordinary pain-killers do cause the later painful states. However, there are theories of causation and mental state identity according to which this is not so, and I don't want my example to depend on the truth of any particular controversial metaphysical view.

¹⁴ The example can be modified to make it closer to Keeper's situation. Simply assume that the drugs also cause Patient's death at about the same time that he would have died anyway.

¹⁵ This involves a backtracking counterfactual.

¹⁶ Alan Donagan, *The Theory of Morality*, p. 42. (Chicago: Chicago University Press, 1977).

¹⁷ Donagan, pp. 42–3. For a useful discussion of this passage and the distinction between doing and allowing contained therein, see Jonathan Bennett, "Negation and Abstention: Two Theories of Allowing", *Ethics*, October 1993.

¹⁸ For example, this analysis seems to suggest that human agency is somehow outside nature, that human agents have the power to break the laws of nature.

¹⁹ The ideas of this section, in particular the first two paragraphs, have been greatly influenced by discussion with and unpublished writings of Jonathan Bennett.

²⁰ See Bernard Williams "A Critique of Utilitarianism", in J. J. C. Smart and Bernard Williams, *Utilitarianism For and Against*, pp. 98ff. (Cambridge: Cambridge University Press, 1973).

²¹ The original example does not provide a dilemma on this account of dilemmas and ill-doing. Jim does not do ill by the one Indian by shooting him. If Jim doesn't shoot him, then Pedro will.

²² This would make an ordering of actions based on the adjusted value of consequences agent-relative. This fact on its own should give a utilitarian, or any consequentialist, reason to be suspicious of this move.

²³ For a detailed discussion of the issues, see Jonathan Bennett, "Morality and Consequences", in S. McMurrin (ed), *The Tanner Lectures on Human Values*, vol. II, pp. 49–116. (Salt Lake: University of Utah Press, 1981).

²⁴ It's not the mere desire, since a very weak desire that others be well off is quite consistent with a thoroughgoing egoism.

²⁵ This was suggested to me in conversation by Steven Lee.

²⁶ It is interesting to note that malevolence may be unlike benevolence in this respect. The truly malevolent person may have a desire to be personally involved in the misfortunes of others. This may even cause a malevolent person to produce less suffering than would otherwise have occurred. I'm not sure that this is the whole story, though. I suspect that the malevolent person desires not just to be personally involved in the suffering of his victim, but that the victim be aware of the involvement. The victim's awareness of his tormentor's role in the production of his suffering may often increase the suffering itself. Recall Feste's revelation to Malvolio of the part he played in Malvolio's humiliation in *Twelfth Night*.

²⁷ I am not sure whether such a modified theory would actually be a version of utilitarianism or a departure from it.

²⁸ This assumes that an act cannot be both right and wrong. Utilitarians are not inclined to challenge this assumption.

²⁹ A deontological theory whose sole command is 'always keep your promises' may allow for moral dilemmas without giving weight to the making/allowing distinction, but it would clearly not be plausible.

Department of Philosophy
Southern Methodist University
Dallas, TX 75275-0142
USA