

Chapter 3

Consequentialism and Motives

In chapter 1 I introduced the thesis that Motives Matter:

MM: There is an action X such that if X were performed from one motive it would fall into one deontic category, and if X were performed from another motive it would fall into a second deontic category in virtue of this difference in motives.

The central tasks of this book are to see if MM is true and, if so, to understand why it is true. In this chapter I will examine the consequentialist position with regard to MM. I conclude that consequentialism can support MM. This is a bit surprising, because consequentialists have sometimes denied MM, either implicitly or explicitly. In examining how consequentialism can support MM I will bring out that this tradition contains two streams of thought bearing on our topic. I will call them extrinsic and intrinsic consequentialism. We will see how they differ in the support they give to MM.

I

A Brief Survey

The consequentialist approach to rightness and the deontic categories could be stated summarily in this way: at a given time an agent is morally obligated to choose the action open to her that has the best consequences. Any action that she chooses that has less than the best consequences is wrong. The utilitarian version of consequentialism holds that at a given time an agent is morally obligated to choose the action open to her

that produces the most happiness. Any action that she chooses that produces less than the most happiness is wrong. These statements leave many questions open about the deontic categories and their application, as well as about the interpretation of terms like ‘consequence’. At certain points we will need to take up some of them, but I propose to postpone doing so. It is clear even from these statements that in consequentialism motives do not figure in the theory at the most basic level. But there is the possibility that they are deontically relevant in some derivative way. It is also clear that in consequentialism deontic status is largely objective. The considerations bearing on the deontic status of an action are objective facts about it, such as whether it brings about injury or health. Notice that I am using ‘objective’ in such a way that facts about an action like producing pleasure or pain count as being objective. Pleasure, pain, happiness, and unhappiness are sometimes said to be subjective states, but for our purposes the fact that an action produces pleasure or pain counts as an objective fact about it.

In this section I survey what the leading consequentialists have said about the deontic significance of motives. My goal is to see whether the theory leads to a completely objective conception of deontic status.

The General Tendencies of Motives I noted in the last chapter two conceptual points that Bentham makes about motives. They concern the evaluative component in many motive words, and the variety of actions that a given motive may lead to. Having made these points, Bentham himself goes on to apply utilitarianism to the moral evaluation of motives. There are different sorts of evaluations that a utilitarian could carry out. One of them involves consideration of the general tendencies of some motives. And this he does.¹

It is clear why some motives generally have better social effects than others. Motives establish ends, and the pursuit of some ends has better effects than the pursuit of others. If we think of motives in terms of desirability characterizations some lie at one extreme and characterize the promotion of the well-being of some or all human beings as desirable. Others lie at the other extreme and characterize the suffering of some or all humans as desirable. (There is in fact a more comprehensive contrast that is possible between characterizing the well-being of some or all sentient beings as desirable, and characterizing the suffering of some or all sentient beings as desirable.) The crucial point is that people acting on these motives will often bring about these effects. Even motives that do not lie at the evaluative extremes—the desire for entertainment, say, or curiosity—can have significant effects on human (or sentient) well-being when considered as general tendencies in society. We can be confident in making general evaluations of some motives, and even in some sort of ranking of them. With other motives—self-interest, for example—it may be difficult to generalize at all: there is too wide a variety of things that people do that fall under this broad term. In his treatment of motives Sidgwick writes that “moralists [of various schools] diverge widely in estimating the ethical value of Self-love.”² Bentham is inclined to regard “physical desire”, “love of power”, “pecuniary interest”, and “Self-preservation” as neutral or indifferent.³ He then produces another sort of ranking that considers only the effects of motives on other members of the community.⁴

The classical utilitarians are all aware that no motive whatsoever has uniformly good or bad effects. The conceptual point we noted in the last chapter means that a motive like sympathy can lead to various specific actions, like making a phone call or

sending a check. These actions can have good or bad consequences. Genuinely sympathetic actions can have terrible consequences if an agent makes false factual assumptions. Think of well-meaning support for lobotomies or compelling Native American children to speak English. The same point applies to a bad motive like revenge. It can lead to actions like ‘whistle blowing’ or ‘informing’ that have good consequences. This being true, Bentham thought that the evaluation of a given action from a motive must consider what its effects are in that case.⁵ He states, “if they [motives] are good or bad, it is only on account of their effects.”⁶ Sidgwick also recognized the variable consequences of a given motive.⁷ Mill, we saw, said that a wrongful act can be done from friendship.⁸

Apparent Support for PRC Some forty years after Principles was published

John Austin wrote as follows:

Of all pleasures bodily or mental, the pleasures of mutual love, cemented by mutual esteem are the most enduring and varied. They therefore contribute largely to swell the sum of the well-being, or they form an important item in the account of human happiness. And, for that reason, the well-wisher of the general good, or the adherent of the principle of utility, must, in that character, consider them with much complacency. But, though he approves of love because it accords with his principle, he is far from maintaining that the general good ought to be the motive of the lover. It was never contended or conceived by a sound, orthodox utilitarian, that the lover should kiss his mistress with an eye to the common weal.⁹

Austin is here focusing on a specific action. He was probably influenced by Adam Smith’s treatment of the ‘invisible hand’, and his claim that socially beneficial actions may result even from ‘self-love’. We are naturally inclined to look at what he is saying in relation to Kant. Austin can then be seen as making two of Kant’s points in the context of utilitarianism. First, a given sort of action may be performed from different motives;

second, an action may conform to duty and yet not be performed from duty. That is, an act that is morally obligatory in utilitarian terms need not be done from a desire to perform an act which is obligatory, or from a desire to produce the most happiness. The principle of utility requires agents to perform actions that produce the most happiness. Since the motive of love here leads to an action that produces the most happiness that action is morally obligatory. If this is correct we might go on to conclude that the action would remain obligatory if it were performed from another motive. If that is correct, Austin could be taken to be employing utilitarianism to endorse Prichard and Ross' Claim (PRC).

PRC: All actions that are morally obligatory are specified without mentioning their motives.

Notice, though, that if this is what Austin is doing, he is not basing his support of PRC on the two other arguments supporting PRC that I mentioned in Chapter 1. The support seems to come from the utilitarian principle itself.

Sidgwick, in a well-known passage, seems to echo Austin's support for PRC.

...the doctrine that Universal Happiness is the ultimate standard must not be understood to imply that Universal Benevolence is the only right or always the best motive of action. For, as we have observed, it is not necessary that the end which gives the criterion of rightness should always be the end at which we consciously aim: and if experience shows that the general happiness will be more satisfactorily obtained if men frequently act from other motives than pure universal philanthropy, it is obvious that these other motives are reasonably to be preferred on Utilitarian principles.¹⁰

Sidgwick is here partly thinking of motives as character traits, since he is considering how often people will perform socially beneficial actions if they have certain

motivational patterns. In other words, he is suggesting that if a person is friendly, say, she will tend to perform more useful actions than if she had other characteristic motivations.¹¹ If so, a utilitarian would encourage people to develop this trait. There is a close connection between character traits and motives that we will consider later. But Sidgwick seems also to be echoing Austin's point about a given action that produces the best results. Such an action need not have been performed from a desire to produce the best results.

In more recent consequentialist writing the contrast that has come into prominence is not the older one of motive versus rightness, but that of decision procedure versus standard (or criterion) or rightness. A number of prominent theorists have argued that consequentialism should be construed only as standard of rightness and that it can allow and even encourage agents to use some other method of moral or practical reasoning in their everyday affairs.¹² This means that their motives can be tied to these forms of reasoning and not to any consequentialist principle. So if, for example, it is socially beneficial for people to cultivate friendships and close relationships then consequentialism can encourage people to focus their deliberation in many cases on what will further these relationships, and to act from the motives that establish the features of these relationships as ends. This point obviously needs to be balanced somehow with Mill's point that a desire to help a friend can lead someone to act wrongly.

II

The Quasi-Millian Argument for the Falsity of MM

MM is not a thesis about the general tendencies of actions. It concerns the significance of motives with regard to the deontic status of specific actions. We have seen that even if PRC is true that MM could be true. So even if there is evidence that consequentialism supports PRC it might still accept MM. Let us consider whether there is a general argument based on consequentialism that leads to the denial of MM.

There is one well-known (but brief) passage in which a utilitarian argues that MM is false. (I will mention a less well-known passage in Moore to this effect below.) This is the passage in Mill's Utilitarianism quoted in Chapter 1.¹³ It suggests that motives are irrelevant with respect to both obligatory and wrong actions. Mill seems to have relied on two lines of thought here. The first is the one suggested by the passage already quoted from the first edition. The second is developed in a long footnote that was added to later editions. Here Mill responds to two counterexamples to the conclusion in the first edition put to him by a critic.¹⁴ I believe the material here is less persuasive, as I explain in a note.¹⁵ In any case, Mill must have continued to think of the material in the first edition as persuasive, since he did not withdraw it, but only added to it. So let us look at it again. The crucial statements are these:

He who saves a fellow creature from drowning does what is morally right, whether his motive be duty or the hope of being paid for his trouble; he who betrays the friend that trusts him is guilty of a crime, even if his object be to serve another friend to whom he is under greater obligations.¹⁶

In chapter 1 I sketched an argument based on these claims. I will set it out again. This argument probably goes beyond what Mill meant.¹⁷ In fact, I acknowledge that I am about to construct a straw man. My purpose is to facilitate the discussion. Let us call it the quasi-Millian argument.

In consequentialism rightness and wrongness depend on the consequences of an action. That act among an agent's options that has the best consequences is right. But motives are not consequences. (Mill himself thinks of a motive as a "feeling."¹⁸) At the time of an action its motive is water that has just passed under the bridge.

Consequentialism makes the deontic status of this action rest on the value of what follows it, as compared to the value of what would have followed the other actions that the agent could have performed. It is a completely 'forward-looking'¹⁹ theory, that is, forward-looking in time. So the betrayal of a friend is wrong because its consequences are worse than remaining faithful to her, and this difference would persist if the betrayal were the product of a good motive. Likewise, the goodness brought about by saving a person from drowning, as compared to leaving her to drown, is what makes saving her obligatory.²⁰ Performing this action from another motive would not alter the difference in consequences. If all of this is correct then consequentialism entails not only a largely objective conception of deontic status, but a completely objective conception of deontic status. And it not only affirms PRC, but denies MM.

III

Problems with the Argument

The quasi-Millian argument for the falsity of MM has defects. In this section I describe two of them.

The Value of the Act Itself. The argument assumes that the deontic status of an action is determined only by the value of its consequences. But this formulation invites a question about the value of the act itself. An act is not a consequence or effect of itself. Nonetheless, consequentialism has to include any intrinsic value that the action itself has in the calculation that determines its deontic status. This is because the choice to do one action rather than another will not only mean that the world will differ with respect to the consequences of the two actions. It will also differ with respect to the fact that one action rather than another is performed. The total difference in the world that an action makes must reckon on the difference the act itself makes. This is a point that is occasionally noted in passing.²¹ But it is important in thinking about MM.

We now need to clarify the notion of consequence in the statement of the consequentialist principle. Let us begin by distinguishing between the narrow consequences of an action and its broad consequences. Narrow consequences include only those events that occur after an act, qua intentional bodily movement. ‘Narrow consequences’ include two kinds of events (or facts): first, the causal effects of an action, considered as an intentional bodily movement; second, any states of the world following the action that would have been different had the agent chosen to act differently. If S shoots and kills T then the death of T is a narrow consequence of S’s action. Likewise, if

S sees T drowning and chooses not to rescue T then the death of T is also a narrow consequence of S's action (that is, her intentional inaction). This sort of conceptual inclusiveness is commonly accepted, I think, by consequentialists. The notion of a broad consequence is more inclusive still. Broad consequences include all narrow consequences, as well as the features of the act itself. The terminology here is my own, but all consequentialists agree that deontic status is determined by the value of the broad consequences of actions. That is to say, they agree that if an action X itself has any intrinsic value then this must be included in the moral calculation that determines if X is right.

The intrinsic value of the act itself has importance even for the simplest version of consequentialism, utilitarianism. After all, one action may be more pleasurable for an agent to perform than another. Such a difference might alter the results of the calculation determining which of these actions has better (broad) consequences. This point could make an act's motive indirectly relevant deontically in utilitarianism. The pleasure an agent receives from performing an action is not its motive, but its motive could influence how pleasant it is to perform that action.²²

The possibility that an act itself can have intrinsic value is important for other versions of consequentialism. Pluralist forms of consequentialism hold that other things besides pleasure and pain have intrinsic value and disvalue. It might be said, for example, that beauty is intrinsically valuable. This also creates the possibility that features of actions themselves are deontically relevant, since actions (in ballet, for example) can be beautiful. Here again the act's motive could be indirectly relevant deontically, since it

might be that performing an action from one motive rather than another would make it more beautiful.

There are various reasons why someone might think that actions, or features of actions, are sometimes intrinsically valuable, or intrinsically bad. For our purposes the most important reason is the idea that their motives themselves are sometimes intrinsically valuable, or intrinsically bad. I will describe just such a form of consequentialist theory below. This is an unusual, non-hedonistic form of consequentialism. It denies Bentham's statement, "if they [motives] are good or bad, it is only on account of their effects."

The quasi-Millian argument seems to limit its attention to the narrow consequences of actions. It focuses on consequences like another person's living or dying. However, all consequentialists must include any intrinsic value that an action has in the calculation that determines its deontic status. Motives could affect the intrinsic value of an action in two different ways. This means that they could affect the deontic status of an action by changing how much intrinsic value the broad consequences of an action have. All consequentialists must therefore deny that the argument shows that their theory requires them to deny MM.

Motives and their Further Effects We can now ask whether the quasi-Millian argument is even successful with respect to the narrow consequences of actions. It is not.

The discussion of the nature of motives in the last chapter helps to show that the psychological and metaphysical assumptions about motives made in the quasi-Millian argument are incorrect. The argument depicts motives as prior to the actions they lead to. But we saw that whenever an agent acts on a motive it guides her activity. In some cases

she may even modify her specific intentions in order to achieve the goal that her motive sets for her. It is therefore better to think of a motive as contemporaneous with an action, rather than as prior to it. The onset of a desire will occur before an agent acts on it, if indeed she does so. But when she acts on it the desire has not receded into the causally inert past. If S wants to hear a concert that starts at 8, for example, she will drive to it in a way that she believes will insure that she arrives at the hall before that time. Bentham, among others, speaks of the effects of motives. Presumably he would agree that this sort of guidance is one of its effects.

The fact that motives guide the performance of actions means that when the same kind of action is performed from two different motives it may be performed in somewhat different ways. This fact plays an important role in detective stories, and, presumably, in real detective work. Someone who kills in revenge will do so in a way that differs from someone killing out of greed. In the former case, for example, the killer may leave the victim's wallet in place. So, if some peculiar lover were actually to "kiss his mistress with an eye to the common weal" he would not do this in quite the same way as if he did it from love. Hence, if S kisses his mistress from affection, and T kisses his mistress "with an eye to the common weal", S's mistress might find the kiss to be more satisfying simply as a physical movement. The point applies especially to those actions that Bernard Williams called "human gestures".²³ So a consequentialist cannot really suppose that all acts are such that they produce the same narrow consequences no matter what motive leads to their performance.²⁴

This point can be made in another way by considering the rational relations of an act and its consequences. Consider a merchant S who gives the correct change to

someone because it is in her self-interest and another merchant T who does this from a sense of duty. Each will be focusing on somewhat different features of the action, and what they do intentionally will necessarily differ. S gives the correct change intentionally and furthers her self-interest intentionally, but she does what is right, at best, only as a foreseen consequence of her action—if she even believes that it is right. In other words, for S doing what is right is a by-product that she does not bring about intentionally. T also gives the correct change intentionally, and does the right thing intentionally, but she furthers her self-interest, at best, only as a foreseen consequence of her action—if she even believes that it does this. For T, the furthering of her self-interest is a by-product that she does not bring about intentionally.²⁵ This is not to say that there will always be a morally relevant difference between the narrow consequences of performing a certain action from two different motives. But it suggests why this is possible.

There is another, partly over-lapping, reason why motives can make a difference in the narrow consequences of an action. Motives can be of interest to another person, even when they lead to perceptually indistinguishable actions. A number of the points in this paragraph and the following one are made in the excellent discussion of Thomas Scanlon, in his forthcoming book, Moral Dimensions.²⁶ He speaks of the ‘meaning’ that an action has for an agent and others; this derives from the agent’s reasons for acting (both ultimate and derived). If T is the recipient of a favor from S then T might be interested in knowing S’s motive, that is, her reason, for so acting. T might have one sort of reaction to S’s action if she believes that S acted from self-interest, and another sort of reaction if she believes that S acted from affection. This important fact about human nature helps to explain why people often try to conceal their motives.

One reason that we care about others' motives is that they are evidence about the further relations that we will have with them, conceived of simply in terms of actions without motives. That is, I may infer from the fact that you acted from a certain motive that you are likely to kick me—or to kiss me—in the future. I may also make further inferences about what motive you will act from. And we care about what use another person will put our interactions to. Scanlon gives the example of S wanting to know why T invites her “to the big end-of the year dance”. S might want to be reassured that T was not merely seeking a chance to associate with the ‘in crowd’.²⁷ These sorts of concern obviously rest in part on a more basic one: to understand an agent's reason for acting on the present occasion. Even if T expects never to meet S again, she may have a different feeling about the favor that S provided her if she believes that S had one motive rather than another.

This point has had a particular significance with regard to the motive of duty. Kant places great emphasis on this motive, and famously claims that only actions stemming from it have moral worth.²⁸ Sidgwick may deserve the credit for first raising doubts about whether the recipients of actions motivated by a sense of duty are always gratified to think that this is what led the agent to act. He writes, “Benefits which spring from affection and are lovingly bestowed are more acceptable to the recipients than those conferred without affection, in the taste of which there is admittedly something harsh and dry.”²⁹ In our own day Bernard Williams³⁰ and Michael Stocker³¹ have echoed this thought, giving well-known examples of agents disappointed to learn that someone helped them from a sense of duty. Lawrence Blum believes that “in general one prefers to be helped from sympathy [rather] than from duty; for the former response conveys a

greater good than does acting from duty.”³² If this is correct then consequentialists can not only take issue with the Kantian claim that the sense of duty is always a valuable motive. They can even assert that sometimes it is wrong-making.

In sum, motives affect how an action is done, and people care, for various reasons, why actions are done. So a consequentialist will not grant that the narrow consequences of an action will be the same, no matter what motive leads to it. Hence, motives can be relevant deontically because of their influence on the narrow consequences of actions. The quasi-Millian argument ignores this point.

We have now seen that any form of consequentialism that accepts Bentham’s view that motives are valuable only because of their effects should say that they can be relevant deontically in three different ways.

1. They can indirectly affect the intrinsic value of the action itself;
2. They can affect how an action is done, and thus indirectly affect the intrinsic value of an act’s narrow consequences.
3. They can influence people’s responses to the act because they care about why the agent did what she did. These responses in turn can affect the intrinsic value of the act’s narrow consequences.

These three points all can be captured by saying that motives are deontically relevant because of their extrinsic value. That is, they all assert that in different ways motives can have effects on how much intrinsic value there is in the world. The theory I mentioned

above, and that I will discuss shortly, accepts the previous three statements, as well as the following one.

4. Some motives are intrinsically good and others are intrinsically bad.

If this statement is true then motives can affect the intrinsic value of an action in another way, and thereby affect the intrinsic value of the act's broad consequences in another way. All consequentialists therefore have three reasons for denying that the theory leads to a completely objective conception of deontic status that rejects MM. Some consequentialists have a fourth reason to deny this. They all should accept MM.

IV

An Example

In this section I present an example that shows a motive that is deontically relevant for consequentialism. I have devised one that shows it to be relevant even for the familiar, Benthamite forms of consequentialism that take motives to be relevant only because of their effects. I propose to examine the motive of racism. In Chapter 1 I discussed an agent's refusal to sell her house to another person because the buyer belongs to a certain ethnic group. Here I will use another, simpler example of this motive.

Let us consider two agents, S and T, who are dignitaries unexpectedly called upon to greet certain people. One of these people is R, who is black. R does not know much about the racial attitudes of S and T, and admires both of them. T's right hand is injured, and is bandaged. If T shakes R's hand it will be painful for her, so the avoidance of pain

is on T's mind. S, on the other hand, is a racist. She has contempt for blacks, and no compunction about showing it.³³ If S refuses to shake R's hand then it will be obvious to both of them that her motive is racism. R will be humiliated and angered by this. We are interested in the deontic status of the refusal to shake R's hand. The numbers below represent units of pleasure and pain. It is admittedly artificial to do this, but it will simplify the relevant points.

<u>S's Options</u>			<u>T's Options</u>		
	Shake	Don't Shake		Shake	Don't Shake
S	-5	5	T	-15	0
R	10	-10	R	10	-2
Total	5	-5		-5	-2

The numbers represent the following psychological facts. R will be gratified to shake both agents' hands, hence the value of 10 for her of doing so. S would be displeased to do this. If S doesn't shake R's hand, she will be pleased. R, on the other hand, will be quite unhappy. R's stake in this case is greater than S's: there is a difference of 20 for her in the two options, but only a difference of 10 for S. It will be physically painful for T if she shakes R's hand, given her injury. Hence shaking has a value of -15 for T. If T does not shake R's hand then R will be slightly disappointed. R will not be deeply upset, since she can see T's bandages and understand T's reason for acting. T will be neither pleased nor displeased by not shaking. In this second case it is T, not R, who has the bigger stake in T's decision.

If we suppose all of this, it follows that S acts wrongly in refusing to shake R's hand and T does not. Part of the reason for this verdict is that the motives of S and T have effects on R's happiness. We thus have an example of this structure: X from M1 is wrong but X from M2 is (broadly) permissible. The motive of racism has deontic relevance for a hedonistic form of consequentialism.

Some comments are in order about the reasoning just presented. First, certain assumptions were made about the effects of open racist motivation and the existence of alternatives. But there is nothing objectionable in making special assumptions, which are in any case not unrealistic, in supporting the claim that motives are relevant deontically. To say that motives are relevant is only to say that they are sometimes relevant. The assumptions can therefore be taken to illustrate when they are relevant.

Second, I used utilitarianism in the example. But it is clear that other forms of consequentialism, employing different assumptions about intrinsic value, can also support MM. This is true even if they do not hold that motives themselves have intrinsic value. We see this if we remind ourselves that often pluralist theories of intrinsic value assert that pleasure is one of the things that is intrinsically good and pain one of the things that is intrinsically bad. But even if a consequentialist theory denied this and held, for example, that dignity is one of the things that is intrinsically good then an example of this form could be constructed. Indeed, Scanlon, who rejects consequentialism altogether, accepts that the meaning of an action can affect its permissibility, and he grants that a racist act such as I described could be wrong because of the effects it has, or would be likely to have, on its target. So the example certainly does not isolate one particular moral

theory that can account for it. But we see how familiar forms of consequentialism can grant that the act is wrong in virtue of its motive.

Third, the sort of case described represents one type of wrongful racist act that constitutes a sort of insult. But many other actions that are wrongfully discriminatory work in a fundamentally different way. Scanlon notes that when discrimination is widespread then members of the target group are commonly denied certain goods and opportunities.

Once a practice of discrimination exists, decisions that deny important goods to members of the group discriminated against and do so without sufficient justification, are wrong even if they express no judgment of inferiority on the agent's part. They are wrong even if done simply out of laziness, or a desire to avoid offending others by going against established custom.³⁴

We might put this important point by saying that a practice of racial discrimination constitutes a sort of conspiracy to deny benefits to members of a particular race. Religious and other forms of discrimination are likewise conspiracies. These discriminatory practices represent a transmutation of the underlying motives. Motives like racism, when pervasive, can bring about the existence of practices governed by norms. These norms are then usually experienced by agents as being analogous to moral norms, or, indeed, as being moral norms strictly speaking. Agents in the relevant group therefore generally regard conformity with the norms as something that they 'have to' do. For example, they believe that they have to refuse to rent their apartments to members of certain races or religions. They may then think of these norms as governed by PRC. If so, they believe that the relevant discriminatory acts can acceptably be performed from various motives. Other agents may reject the norms, but deny goods or opportunities to

the victimized race for self-interested reasons. When this situation exists, what we might call ‘objectively racist’ actions, for example, will occur that do not always stem from racist motives in the strict sense. The moral evaluation of such actions need not make any reference to the motives of the agents. As Scanlon suggests, the wrongness of these discriminatory actions is due to the denial of benefits and opportunities to the victims. Participating in the conspiracy is now an objective wrong, like participating in a criminal enterprise. (Issues of blame pose many further questions: for example, how much pressure exists on agents to conform.) Granting all of this does not undermine the force of our example. But it does show that a term like ‘racism’ does not always refer to a motive, but often to a practice that has to be analyzed morally in a different way. It also shows in another way that we have a largely objective notion of deontic status.

Fourth, note that the example supposes there are two agents with different motivational structures. We did not suppose that one agent can refuse to shake hands from racism or from self-interest. To say that motives are deontically relevant need only mean that there are tokens of the type of action ‘refusing to shake hands’ that are wrong in virtue of having one motive and other tokens of that type that are permissible in virtue of having another motive. It is a separate question whether it is ever the case that one agent has both of these tokens available to her in a given situation. I will address that question in Chapter 7, but it need not be taken up yet.

The example supports the conclusion that familiar forms of consequentialism, which assume that motives are deontically significant only because of their effects, support MM. This conclusion is appealing when we recall a question that I posed in Chapter 1. I noted that a number of philosophers have claimed that motives are relevant

to various moral judgments like those concerning character or blame. But they hold that they are not relevant deontically. A proponent of such a view would need to explain why motives are relevant to other moral judgments and not to deontic judgments. We now see that consequentialism can avoid this difficulty. If motives sometimes make a difference to certain of the consequences of actions then MM is true. Motives thereby have a derivative relevance deontically. Of course, motives can also be relevant to other moral judgments.

V

Intrinsic Consequentialism

Bentham asserted that motives have value only in virtue of their effects. Let us call any form of consequentialism that accepts this assertion ‘extrinsic consequentialism’. I mentioned that there is a form of consequentialism that attributes intrinsic value to motives themselves. In this section I describe such a theory and how it differs from the more familiar forms of extrinsic consequentialism. It can accept that MM is true partly because of a different interpretation of the phrase ‘in virtue of’.

G. E. Moore had the materials at hand to construct this other form of consequentialism, but failed to do so. In Principia Ethica (1903) Moore defends the view that morally obligatory actions produce the greatest amount of good.³⁵ He also claims that certain attitudes have intrinsic value. He holds that “a love of some intrinsically good consequence which he [the agent] expects to produce by his actions” is intrinsically good, as is “a hatred of some intrinsically evil consequence he hopes to prevent” by his action.

“The emotion excited by rightness as such” is also intrinsically good.³⁶ On the other hand, the love of evil and the hatred of what is good are bad in themselves.³⁷ Presumably these desires and emotions can be motives. And Moore is alive to the abstract possibility that actions can have intrinsic value, since he mentions it.³⁸ He thus leaves room for saying that the intrinsic value of a motive can change the intrinsic value of an action, and that the intrinsic value of an action can change the intrinsic value of its broad effects. He therefore could have developed what we can call ‘intrinsic consequentialism’.

However, in Ethics (1912) Moore explicitly denies that motives are relevant deontically. He devotes some space to the question and concludes that “right and wrong depend solely on [narrow] consequences” and not at all on motives. He claims that other sorts of moral judgment are affected by an agent’s motives, in particular whether a person deserves blame or praise.³⁹ It is worth noting that we now have seen that two prominent consequentialists, Mill and Moore, explicitly deny MM, along with two prominent deontologists, Prichard and Ross.

Thomas Hurka has recently defended a theory that claims that some motives are intrinsically good, and others are intrinsically bad.⁴⁰ That is, he defends claim 4 above. Hurka says that consequentialism has tended to focus on the causal properties of psychological states like character traits and has neglected their intentional characteristics.⁴¹ His “recursive theory of value”, in contrast, is based on the intentional structure of these states. The main topic of Hurka’s book is virtue and vice, that is, morally significant character traits like compassion and courage. But the sorts of evaluative structures that Hurka finds in character traits are also found in desires. He

therefore asserts that desires can be intrinsically good or bad in virtue of their intentional structure.⁴² This means that motives can be intrinsically good or bad.⁴³

I will give a simplified account of the recursive theory.⁴⁴ It has a series of levels. The base level consists of states of affairs having intrinsic value and badness. Hurka says that pleasure, knowledge and achievement have intrinsic value; pain, false belief and failure are intrinsically bad. At a second level certain psychological attitudes about the things that have these values themselves have intrinsic value or disvalue. This is Hurka's (and Moore's) fundamental and powerful point: there are appropriate and inappropriate attitudes to take to the basic intrinsic values and disvalues, or to the things that have intrinsic value or disvalue. Appropriate attitudes to basic values are themselves intrinsically good; inappropriate attitudes to basic values are themselves intrinsically bad. For example, S can desire to produce pain for T as an end. Hurka is claiming that this desire is itself undesirable or bad, independently of its effects. Of course, his theory will not deny that this desire is also bad in many cases because of its effects. Intrinsic consequentialism can accept everything that extrinsic consequentialism says. It simply adds to the claims made by extrinsic consequentialism. Intrinsic consequentialism therefore asserts that motives are deontically relevant 'in virtue' of their effects, and 'in virtue' of their nature as intentional states.

There could be other versions of recursive theory with different views about which desires have intrinsic value and disvalue, and why. A different version could stipulate other base-level intrinsic values, such as beauty. For our purposes, though, the important points about Hurka's theory are the following. Desires can be appropriately or inappropriately oriented towards the basic things that have intrinsic value and disvalue. In

virtue of these orientations desires themselves can have intrinsic value. These claims can be accommodated within a consequentialist approach to deontic matters. The sort of consequentialism we get from his theory can still say that the right action is the one that has the best broad effects. The only difference between the form of consequentialism that we get by adding Hurka's claims to extrinsic consequentialism is that the intrinsic value (or badness) of a motive is factored into the calculation of the action's broad consequences. This value may be small in comparison to the narrow consequences of the action. It follows that an action from an intrinsically bad motive could be right, and an action from an intrinsically good motive could be wrong.

I postpone a more detailed discussion of Hurka's intrinsic consequentialism until chapter 8. But I will mention here one issue that seems to speak in its favor. This is the issue of concealed motives. We can imagine a case where a seller of a house refuses to allow a black person to buy it, and manages to conceal her racist motivation. In this case there will not be the same strong negative reactions by the buyer, although there may be muted ones fueled by suspicions about the seller's motive. We still are inclined to think that refusing to sell from this motive is wrong. Extrinsic consequentialism seems to have a difficult time endorsing this assertion. Its proponents can, of course, appeal to the further effects of the motive on the agent's character, and hence indirectly on others. But if the agent always tries to conceal her racism, or acts only occasionally on open racism, these further effects may not be significant enough to show that a refusal to sell now from concealed racism is wrong.

Intrinsic consequentialism does not seem to have the same difficulties in dealing with concealed motives. One of Hurka's assertions is that a malicious motive is

intrinsically bad. Even concealed malice is a real feature of the world: something bad in itself but concealed is still bad. Therefore even a refusal to sell from concealed malice could be wrong because of the change that its intrinsic badness makes in the calculation of the total intrinsic value of the broad consequences of the action. Hurka tends to focus on abstract value structures like malice rather than more specific ones like white racism or anti-Semitism. It is reasonable to assume, though, that his theory can show that these forms of motive are also intrinsically bad. We can therefore see how even an action stemming from a concealed racist motive could be wrong in virtue of its motive.

Intrinsic consequentialism is a rare form of the theory, and Hurka is the only writer I know of who has developed the materials needed for it. We do not need to explore it fully at this point because it only adds a reason to those endorsed by all consequentialists for saying that MM is true.

VI

Conclusion

Consequentialism contains two different positions on the value of motives, intrinsic and extrinsic consequentialism. The main stream of the theory, beginning with Bentham, should be thought of as endorsing extrinsic consequentialism. Extrinsic consequentialism actually supports MM, since motives can alter the intrinsic value of an action, as well as the intrinsic value of its narrow consequences. Intrinsic consequentialism also supports MM.

The discussion thus far leaves a number of questions open. We have only looked at one of the three sorts of example presented in Chapter 1. It is not obvious that the points playing a role in my example can be appealed to in the example of prostitution. We easily see the evil of racism's effects. And we have some sympathy for the suggestion that the basic evaluation in some forms of racism is intrinsically bad, in a way that Hurka's approach to psychological attitudes can probably explain. So my example involved adding a bad motive (extrinsically and possibly intrinsically) to what might seem like an otherwise neutral act. But the suggestion that the desire for money is wrong-making when it leads to having sex is different. A desire for money moves all of us to do many things, and it does not seem to be a bad motive. Bentham regarded it as neutral. And an act of consensual sex does not strike us as problematic in itself in the way that killing or lying does. Finally, if T willingly gives S money to have sex, T is unlikely to be upset by knowing S's motive. So if S engages in consensual sex for money, why is her act wrong (if it is)? A "mysterious piece of moral alchemy"⁴⁵ seems to be going on where, it seems, a neutral motive plus a neutral (or even good) act yields wrongness. A consequentialist treatment of the wrongness of prostitution thus requires separate treatment.

A second open question is the existence of obligation-making motives. The example I chose of a deontically-relevant motive was one that is wrong-making. I noted in Chapter 1 that many of the intuitively-convincing examples of motives that are deontically relevant are wrong-making. Both forms of consequentialism can show why motives are sometimes wrong-making. If we left the discussion there we might simply add that consequentialism also supports PRC. Consequentialism would then be saying

that motives are sometimes wrong-making but never obligation-making. In other words, MM is true, and PRC is also true. Consequentialists would have no need to reject the two Kantian-inspired arguments presented in Chapter 1 that also support PRC. This is a difficult position for them to maintain, however. The reasoning that leads consequentialism to the conclusion that motives are sometimes wrong-making also leads it to the conclusion that they are sometimes obligation-making. If extrinsic consequentialists say that motives can have bad effects that thereby make an action wrong that otherwise would not be, it seems that they should say that there are motives whose good effects make an otherwise wrong action obligatory. And a similar point applies to that part of intrinsic consequentialism that concerns the intrinsic values of motives. Intrinsic consequentialists assert that motives can be intrinsically good and bad. This point should also tell in both directions: the intrinsic value of some motives should make an otherwise wrong action obligatory. What this means is that all consequentialists must confront the two arguments that support PRC. Since extrinsic and intrinsic consequentialism both apparently lead to the conclusion that motives are sometimes obligation-making, proponents of these theories need to explain why these two arguments are mistaken. These matters will be taken up in Chapter 8. But we do not need to address them now. If we can see why consequentialism supports the claim that there are wrong-making motives we have enough to proceed.

Finally, I have not tried to settle the question of whether some form of extrinsic or intrinsic consequentialism is more defensible. What we have got at this point is some sense of the potential of consequentialism to support MM. This speaks in favor of consequentialism. I propose to consider next some other moral theories and how they

relate to MM. After we look at them it will be necessary to return to the evaluation of intrinsic and extrinsic consequentialism.

¹ IPML 115f. Cp. the critical remarks Sidgwick makes about the ranking of motives on a very different basis produced by Martineau. ME Bk. III, ch. 12.

² ME 366.

³ IPML 115.

⁴ IPML 116f.

⁵ IPML 114.

⁶ IPML 100.

⁷ ME 369-71.

⁸ U 18.

⁹ John Austin, The Province of Jurisprudence Determined, ed. H. L. A. Hart (New York: Noonday, 1954), pp. 107-8. (First published, 1832.) Note that on pp. 108-11 Austin proceeds to withdraw his assertion in part. Compare the “philosophical joke” put into Bentham’s mouth by John Rawls: “When I run to the other wicket after my partner has struck a good ball I do so because it is best on the whole.” John Rawls, “Two Concepts of Rules,” Philosophical Review 64 (1955), p. 27, n. 24.

¹⁰ ME 413 (cf. 202-10, esp. 202-4); cp. J. J. C. Smart in Utilitarianism: For and Against (Cambridge: Cambridge University Press, 1973), p. 44.

¹¹ This is a point in effect noted by Derek Parfit, Reasons and Persons (Oxford: Oxford University Press, 1984), p. 34 and Peter Railton, “Alienation, Consequentialism, and the Demands of Morality” in his Facts, Values, and Norms (Cambridge: Cambridge University Press, 2003), pp. 151-86, at 170.

¹² R.M. Hare, Moral Thinking (Oxford: Oxford University Press, 1981), chs. 2, 3; Parfit, Reasons, op. cit., pp. 24-43; Brink, Moral Realism, op. cit., pp. 216-17, 256-62; Railton, “Alienation”, op. cit.

¹³ U 18.

¹⁴ U 18, n. 2. For two efforts to explain Mill’s thinking here see Jonathan Dancy, “Mill’s Puzzling Footnote,” Utilitas 12 (2000), pp. 219-22; Michael Ridge, “Mill’s Intentions and Motives,” Utilitas 14 (2002), pp. 54-70.

¹⁵ Much of Mill’s response relies on the distinction he makes between motive and intention. We have already looked at this issue. I think our discussion suggests at least two serious problems in Mill’s treatment of these concepts.

Mill thinks of motives as “feelings”. (Cp. Ryle, Concept, op. cit., ch. 4, who largely treats motives as emotions, in spite of some remarks to the contrary (e.g., p. 88).) The idea that motives are emotions is incorrect in two ways. There are motives that are not emotions, such as self-interest (often) or the sense of duty. And there are emotions that are not motives, like grief. But even if we confine ourselves to emotional motives Mill’s suggestion is unsatisfactory. Mill seems to think of emotions as simply brute feelings, with no propositional content. Cp. Ridge, op. cit., p. 58.

He also does not consider the structure of intentions, and speaks of them generically. If he had considered this he might have seen that there are derived and ultimate intentions. Then he might have seen that there is a case to be made for

identifying ultimate present intentions and motives. We saw that this is not entirely correct, but it is close. Given their closeness it would be difficult to see how they could differ markedly in their deontic relevance. But Mill claims “the morality of the action depends entirely upon the intention”, and never on its motive.

Now Mill does add this qualifying phrase about the irrelevance of a motive: “if it makes no difference in the act.” U 18, n. 2. This implies that a motive can be relevant if it does make a difference “in the act”. This might allow for some of the points made in Section III below. But it is surprising that Mill did not write, ‘if it makes no difference to the intrinsic value of the act or its consequences’, or ‘if it makes no difference in how happy the agent is in acting, or in how much happiness his act produces.’ These phrases pinpoint which differences we would expect Mill to find significant.

Finally, the alleged counterexample that Mill considers involves a compound action, namely, saving a drowning person in order to torture him. This raises special problems having nothing to do with the significance of motives as such. An excellent discussion of this issue is in Scanlon, Moral Dimensions, *op. cit.*, ch. 2, sec. 2.

¹⁶ U 18.

¹⁷ The closest thing to it that I know of is Driver, Uneasy, *op. cit.* pp. 68-70.

¹⁸ U 18, n. 2.

¹⁹ This apt use of the phrase seems to come from H. L. A. Hart, Punishment and Responsibility (New York: Oxford University Press, 1968), pp. 232, 233.

²⁰ Let us add: setting aside the variant of the case discussed in Mill’s note, *viz.*, saving in order to torture.

²¹ Moore, Principia, *op. cit.*, p. 25; Samuel Scheffler, The Rejection of Consequentialism (Oxford: Oxford University Press, 1982), p. 1 n. 2; John Broome, Weighing Goods (Oxford: Basil Blackwell, 1991), p. 4. Cp. Judith Thomson, Goodness and Advice (Princeton: Princeton University Press, 2001), p. 11 n. 3. Robert Nozick notes this possibility for decision theory. Robert Nozick, The Nature of Rationality (Princeton: Princeton University Press, 1993), p. 133.

²² Robert M. Adams, “Motive Utilitarianism,” Journal of Philosophy 73 (1976), pp. 467-81, at 470-1.

²³ Bernard Williams, “Morality and the Emotions,” in his Problems of the Self (Cambridge: Cambridge University Press, 1973), p. 227. Williams used this idea to criticize Kantian, not consequentialist, moral theorists. However, Michael Stocker, “The Schizophrenia of Modern Ethical Theory,” Journal of Philosophy 73 (1976), 453-66, argues in a related vein that both a pure consequentialist and a pure Kantian would not be able to act from the motives that are central to relationships like friendship.

²⁴ This conclusion is certainly endorsed by Brink, *op. cit.*, Parfit, *op. cit.*, and Railton, *op. cit.* All three are concerned to rebut Williams and Stocker. But I have not found a passage in them where they make the point that different tokens of a type of action can have different effects because of their motives. It is probably an unstated assumption. See also Scanlon, Moral Dimensions, *op. cit.*, ch. 1 sec. 5. Contrast Robert Audi, The Good in the Right (Princeton: Princeton University Press, 2004), p. 180.

²⁵ For more on ‘intentional’ see Alfred Mele and Steven Sverdlik, “Intention, Intentional Action, and Moral Responsibility,” Philosophical Studies 82 (1996), pp. 265-87.

²⁶ Op. cit., chs. 1-3. Before seeing Scanlon's book, I wrote a version of this chapter which largely agreed with him on the topics we both covered. (He spends a good deal of time considering the Principle of Double Effect and the Kantian prohibition on using people simply as means, which do not concern the significance of motives.) But I have greatly profited from his work and I have tried to acknowledge my many debts to him.

²⁷ Scanlon, Moral Dimensions, op. cit., ch. 3, sec. 4.

²⁸ G 58 (viii); 65-9 (397-401).

²⁹ ME 223.

³⁰ Williams, "Morality," op. cit., p. 227; "Persons, Character, and Morality," in Williams, Moral Luck, op. cit., pp. 17-19.

³¹ Stocker, "Schizophrenia," op. cit., p. 462. Cp. Michael Stocker and Elizabeth Hegeman, Valuing Emotions (New York: Cambridge University Press, 1996), p. 158.

³² Blum, op. cit., p. 168. See all of his ch. 7.

³³ Jon Elster makes some insightful remarks about 'shaming' actions and their effects in Alchemies, op. cit., pp. 142-3, 146. But he limits himself to cases of a response by S to the violation of a 'social norm' by T.

³⁴ Scanlon, Moral Dimensions, op. cit., ch. 2 sec.8.

³⁵ Moore, Principia, op. cit., pp. 25; 106; 147.

³⁶ Ibid. pp. 177, 179.

³⁷ Ibid., pp. 208-211.

³⁸ Ibid., p. 25.

³⁹ G.E. Moore, Ethics (New York: Holt, 1912), pp. 182-90; in the paperback reprint (Oxford: Oxford University Press, 1965), pp. 77-80.

⁴⁰ Thomas Hurka, Virtue, Vice and Value (Oxford: Oxford University Press, 2001). Hereafter abbreviated VVV. Hurka notes his debt to Moore, VVV 25-6.

⁴¹ VVV 32.

⁴² VVV 42. Hurka briefly notes the deontic significance of such desires at pp. 47-9.

⁴³ A consequentialist actually need not claim that motives have intrinsic value in order to achieve the result that Hurka's theory does. A principle of 'organic unities' makes it possible for a motive with no intrinsic value or disvalue to alter the total value of its an action's broad effects. For this principle see Moore, Principia, op. cit., pp. 27-36. The operation of this principle could make the total value of the unity consisting of an act and its narrow effects different from that of a differently motivated act with identical narrow effects. I will ignore this possibility in what follows.

⁴⁴ For the basics, see VVV 11-23.

⁴⁵ Hart used this striking phrase to describe the contention of retributivism that "the combination of the two evils of moral wickedness and suffering are transmuted into good." Punishment, op. cit., pp. 234-5.