

## Chapter 4

### Testing Kantian Maxims

A common way of contrasting consequentialism and deontology is to say that deontology makes intrinsic features of an action relevant deontically while consequentialism focuses only on the effects of an action. If, in order to pursue this contrast, we assume that motives are intrinsic features of actions then we have seen that this way of putting the difference between deontology and consequentialism is misleading in at least two ways. Prichard and Ross, two noted deontologists, deny MM, and MM is presumably one important way of affirming that intrinsic features of actions are relevant deontically.<sup>1</sup>

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.

On the other hand, a consequentialist like Hurka asserts that the motive of an action can be deontically relevant in virtue of its intrinsic value.

Nonetheless, many philosophers contend that the Kantian form of deontology is distinctive precisely because of the way it focuses moral evaluation on the rational structure of action. Barbara Herman, a distinguished proponent, writes, “Kantian theory neither describes nor judges actions except as they are willed.”<sup>2</sup> But as far as motives are concerned, our survey in Chapter 1 suggested that Kantianism also entails that MM is false. It makes three claims that seem to support this conclusion: 1) There is a distinction

between acting in conformity with duty and acting from duty; 2) Emotions like sympathy may not be available when an obligation to act arises; 3) A hypothetical moral imperative is impossible. These claims at least lend support to Prichard and Ross's Claim.

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

PRC embodies what I called a completely objective conception of moral obligation. PRC does not entail that MM is false. It is possible that there are no obligation-making motives, but that there are wrong-making motives. We need to look elsewhere in Kantian theory to see if it gives any support to the idea that there are wrong-making motives. There is, in fact, important material in the theory that I have not mentioned yet.

In the present chapter we will look closely at some of that material. We will examine the central feature of Kantian moral theory, the Categorical Imperative. The Categorical Imperative is Kant's term for the supreme principle of morality. This principle is usually understood to be a test that establishes the deontic status of all actions.<sup>3</sup> That is, by utilizing this test we determine if an action is right, wrong or permissible. So if an intrinsic feature of actions plays a role in this test then Kantianism takes it to be relevant deontically. Some versions, or 'formulations', of the Categorical Imperative refer to the 'maxim' of an action. This notion of maxim will be of interest to us in this chapter. A maxim seems to incorporate a reference to the agent's motive for performing an action, and maxims are what Kantianism tests for moral permissibility. So the Categorical Imperative test of maxims might show that MM is true.

We will examine how Kantianism tests maxims for permissibility. We need to consider what maxims are, what role motives play in them, and what the results of the

test entail about the deontic significance of motives. I try to steer the discussion as directly as I can towards a comparison of the Kantian and consequentialist treatment of the deontic relevance of motives. We will see that this is not easy to do. One reason is that the Kantian test of maxims requires us to formulate motives in a special way if there is to be any hope of getting plausible results from it. Racist motives, for example, must become more abstract and ideological, in a sense, if they are to be incorporated into testable maxims. When we eventually find a workable maxim to test for the permissibility of acting in a certain way from a racist motive we see that Kantianism yields the judgment that the action is permissible. This surprising result tells against the plausibility of the theory.

## I

### Motives, Maxims, and Their Logic

We begin with a discussion of maxims. I will develop some ideas about maxims that help to clarify general concepts like that of a wrong-making motive. These ideas will also help us to see how Kantianism can build up a largely objective conception of deontic status.

The Categorical Imperative is the Kantian expression for the fundamental principle of morality. But it is well-known that in the Groundwork Kant states this principle in three distinct ways, generally called ‘formulations’. Furthermore, careful scholars have noted that there are variants of each of the three formulations. And in other works there are still others. Richard Dean, a writer sympathetic to Kant, says that “how

many formulations of the Categorical Imperative there actually are” is “surprisingly contentious.”<sup>4</sup> A survey by Allen Wood might be taken as concluding that there are five formulations in Kant’s writings.<sup>5</sup> There are also scholarly debates about what the relations are among the main formulations. Although it is often said that Kant holds them to be ‘equivalent’, this has been denied.<sup>6</sup> Furthermore, scholars disagree about how to interpret each of the formulations. This means that one scholar may believe that one formulation gives a certain result about the deontic status of an action, and another scholar disagrees. There is another complication concerning two different types of duty in Kantian theory.<sup>7</sup>

Obviously we cannot explore all these issues. What I propose to do in this chapter is examine how the first (or Universal Law) formulation of the Categorical Imperative tests maxims. This formulation has a natural priority for our purposes, since Kant repeatedly employs the term ‘maxim’ in stating it. He does not always use the term ‘maxim’ in presenting the other formulations. One passage in the Groundwork states the Categorical Imperative thus:

The Formula of Universal Law (FUL): “I ought never to act except in such a way that I can also will that my maxim should become a universal law.”<sup>8</sup>

Another, slightly different, wording is:

The Formula of the Law of Nature (FLN): “Act as if the maxim of your action were to become through your will a universal law of nature.”<sup>9</sup>

And immediately after this passage Kant gives one of the clearest examples of a maxim in the entire work: “From self-love I make it my principle to shorten my life if its

continuance threatens more evil than it promises pleasure.”<sup>10</sup> (I alluded to this maxim favoring suicide in Chapter 2.)

A maxim clearly incorporates some sort of description of an action that an agent contemplates performing. The idea that maxims are to be tested for permissibility is in fact placed by Kant in the context of practical deliberation. We are to picture an agent trying to decide what to do, and who is concerned to know if an action she finds it desirable to perform is morally wrong.<sup>11</sup> If she tests her maxim she can know the answer to this question. What sort of information about a proposed action does the maxim contain? Herman and Christine Korsgaard, among other authorities, can be cited in support of the assertion that maxims incorporate a reference to an action’s motive.<sup>12</sup> Korsgaard writes, “your maxim must contain your reason for action; it must say what you are going to do, and why.”<sup>13</sup> Herman writes, similarly, that a maxim states “what the agent is moved to do and for what reason.”<sup>14</sup> Or again: “Maxims must include, in addition to description of the action...the motive from which the agent acts and which she takes to justify her acting...”<sup>15</sup> Note that Herman and Korsgaard do not equate a maxim and a motive (or reason). Rather, they say that a maxim includes a reference to the motive (or reason), but also includes a reference to the action that the motive favors.

These remarks suggest that a maxim is simply an agent’s tentative practical syllogism. Or at least it is a truncated form of it, containing its major premise and its conclusion. (I assume that Korsgaard and Herman will grant that it is an agent’s ultimate normative reason that is included in the maxim.) The Categorical Imperative can then be seen as simply a test of the moral acceptability of the practical reasoning that an agent is prepared to act on. Or, we could say that it is a test of whether acting for a certain reason

is morally permissible. This is exactly the sort of framework that would facilitate discussion of MM.

Let us now consider the sorts of inferences about deontic status that a test of the permissibility of maxims allows us to make. The basic features of the test are dictated by deontic logic. In explaining this I will provisionally make some simplifications. My presentation is unfortunately somewhat abstract. It also will turn out to be incomplete, as I explain in section III.

If a maxim favoring doing X fails the test, then acting on that maxim is morally wrong. If a maxim favoring doing X passes the test, then acting on it is permissible (in some sense). Establishing that X is obligatory is slightly more complicated; it requires two tests. Suppose that the maxim favoring X passes the test. Then the maxim of the omission of the act ( $-X$ ) must be tested. If the maxim of  $-X$  fails the test, then the maxim of X favors an obligatory act. (X is then broadly permissible, but not ‘merely permissible’.) If, instead, the maxim favoring  $-X$  passes the test, then X and  $-X$  are both merely permissible. The agent may do either one.<sup>16</sup> I will assume that it is not possible for both X and  $-X$  to fail the test: that would mean that the agent is in a moral dilemma.

If Kantians seek to test maxims for permissibility we can begin to see how they might claim that features of the action itself are deontically relevant. And we can see (in outline, anyway) how they can affirm MM. If we ever find that a maxim of the form ‘X from M1’ passes the Categorical Imperative test while ‘X from M2’ fails the test (or vice versa) then it is the motive of X is making a deontic difference. In interpreting the schema ‘X from M1’ we need to suppose that ‘X’ stands for a type of action, and ‘M1’ for a kind of motive (like sympathy or self-interest). If ‘X from M1’ fails the test then any

specific instance of X performed from an instance of M1 is wrong. One would think that such a result would have to occur on some occasions; otherwise the inclusion of the agent's reason in the maxim is otiose. If the reason for performing an action is never sufficient to change the deontic category that the act is in, it is hard to see why it is included in the maxim. We might as well test 'X' itself for permissibility, and omit the reference to its motive or reason. I will call the schema 'X from M' the simple form of maxim. It is the same form that I used to represent an action from a motive.

I return to the simplifications I just mentioned. Actually, the most that we could conclude from 'X from M1' passing the test and '-X from M2' failing the test is, precisely, that it is permissible to X from M1 and not permissible to -X from M2. The deontic status of the act types X and -X cannot be established from these two tests alone. We say (and think) that, for example, lying is wrong. And, indeed, Kant is often thought to be a 'rigorist' (or 'absolutist') philosopher committed to asserting this very statement. He did, after all, once argue that it is wrong to lie to an intending murderer about the location of the person she plans to kill.<sup>17</sup> But to make the statement, 'lying is wrong' is— from the point of view of a permissibility test of maxims—actually to make a statement about every maxim favoring the act of lying. It is to say this:

For every M, 'lying from M' fails the test of permissibility.

Most contemporary Kantians believe, however, that the rigorism (or absolutism) of the theory can and should be avoided.<sup>18</sup> They accept that the theory should grant that it is permissible in some cases for an agent to tell a lie, for example. This brings up the second simplification. Instead of conceiving of maxims as having the form 'X from M',

we can think of them as having the form ‘Xing in C from M’. I will call this the conditionalized form of maxim. It incorporates a description of the conditions for acting into the maxim. Kant’s sample maxim about suicide, for example, contains the condition “if its [i.e., life’s] continuance threatens more evil than it promises pleasure”. If lying is sometimes permissible then the following statement would be true:

For some C and for some M, ‘lying in C from M’ passes the test of permissibility.

But, again, if lying in condition C1 is wrong then that entails

For every M, ‘lying in C1 from M’ fails the test of permissibility.

Obviously this revision of the form of maxims still allows for the possibility of an absolute prohibitory rule. This would mean that the following form of statement is true:

For every M and every C, ‘Xing in C from M’ fails the test of permissibility.

This revision also allows for the possibility of absolute moral obligations.

A Kantian will surely want to distinguish between the verdicts that the test could give in principle and those that an agent engaged in deliberation needs to arrive at. A deliberating agent would presumably only test a maxim that she is sincerely inclined to act on, if she hadn’t already tested it. This means, for one thing, that she would typically not test a maxim of -Xing if she had determined that her maxim favoring Xing passed the test. If she is inclined to X, and her maxim favoring Xing passes, she would normally and rightly proceed straightaway to X for the reason specified in the maxim.

Kant is one source of the idea that an agent can perform an obligatory action from various motives or, as we say, do the right thing for the wrong reason. In Kantian terminology this invokes the distinction between acting in conformity with duty and acting from duty. When we apply this distinction to the testing of maxims we begin to see how Kant's theory could yield a largely objective conception of deontic status. A maxim of telling the truth from self interest, for example, should pass the test. And, in general, self-interested maxims that favor actions like keeping promises, not harming others, and providing them with help on certain occasions should pass the test. Maxims of keeping promises from other motives like spite or curiosity should also pass the test. This simple and crucial point has not been emphasized. A lot of effort has been expended on showing why a maxim of lying from self interest, for example, will fail the test. But I cannot recall seeing a discussion by any Kantian of how a maxim of, say, truth-telling from self-interest will pass it.<sup>19</sup> Equally, if there is an action such that every maxim favoring it fails the test, then we can call it objectively wrong.

We saw that three Kantian claims suggest that PRC is true. How is this to be made consistent with the idea that the test of maxims can show that MM is true? There are two answers a Kantian could give. We have seen one of them already: Kant could assert PRC and MM.<sup>20</sup> This is to say that there are wrong-making motives, but that obligation is a completely objective concept. On the other hand, Kantians might want to affirm MM and deny PRC. If PRC is false, then, of course, more would need to be said about how the three Kantian claims are to be interpreted. That is, the Kantian would need to explain why these claims do not entail PRC, as they seem to do.

We now should ask what these assertions about motives mean in terms of the results of testing maxims. Let us first consider what a wrong-making motive would be. Using the schematism for maxims we can now see that this expression is ambiguous. There is a weaker and a stronger concept of a wrong-making motive. In order to understand these concepts we first need to consider the concept of objective wrongness. In explaining all these concepts I will utilize the simple form of maxim, which is all that we need. An action X1 is objectively wrong if a statement of the following form is true.

For every M, 'X1 from M' fails the permissibility test.

The idea of testing maxims as such does not tell us whether any actions are objectively wrong. But it leaves room for that possibility. (And we have a general reason for thinking that deontic status is largely objective, so this result is theoretically attractive.)

The weaker notion of a wrong-making motive might also be described as saying that a motive is wrong-making with respect to a given action. M1 is a unique weakly wrong-making motive if a statement of this form is true:

There is an action X1 that is not objectively wrong, such that there are motives M1...Mn where 'X1 from M1' fails the permissibility test, and 'X1 from M2', 'X1 from M3', ..., 'X1 from Mn', pass the test.

It is possible that more than one motive favoring X1 fails the test, while most motives favoring it pass the test. In that case it still makes sense to speak of the members of the smaller set as being weakly wrong-making. The idea of a strongly wrong-making motive M1 is captured in a statement of the form:

There is a motive M1 such that for every X which is not objectively wrong, 'X from M1' is weakly wrong-making.

These definitions can be explained as follows. If there is an objectively wrong action then the motive from which it is performed is irrelevant: it is wrong in any case. A strongly wrong-making motive means that for any action which is not objectively wrong this motive makes the action wrong. When this is true the action can be performed permissibly from other motives. A motive could thus be strongly wrong-making in two different ways. The first is if there are objectively wrong actions; the second is if there are not. In either case, though, every action that a strongly wrong-making motive favors is wrong, since it makes performing it wrong if it is not already objectively wrong. If there are strongly wrong-making motives there is a sense in which mention of the action they favor in the maxim is otiose. Since doing anything from such a motive is wrong, the agent needn't bother to ask what action the motive favors. This is a sort of rigorism or absolutism about some motives. In contrast, with a motive that is weakly wrong-making the agent would need to specify what action it favors. She might be permitted to perform some actions from this motive. MM is true if there are wrong-making motives in either sense. Nonetheless, we see that these are significantly different possibilities.

Now we can consider what the idea of an obligation-making motive means in terms of testing maxims. There is again a strong and a weak version of this expression. In explaining this we need to keep in mind that obligations get established by two tests on maxims. Let us assume that  $\neg X1$  has failed the test. Then  $X1$  is objectively right or obligatory when a statement of this form is true.

For every  $M$ , ' $X1$  from  $M$ ' passes the permissibility test.

To say that X1 is objectively right is to say that PRC holds of it; it is obligatory no matter what motive the agent performs it from. To say that M1 is a unique weakly obligation-making motive is to say something of this form:

There is an action X1 that is not objectively obligatory, such that there are motives M1...Mn where 'X1 from M1' passes the permissibility test, and 'X1 from M2', 'X1 from M3', ..., 'X1 from Mn', fail the test.

To say this, given the assumption about  $\neg X1$ , is to say that the agent must do X1 and must do it from M1. It is possible that more than one motive favoring X1 passes the test, while most motives favoring it fail the test. In that case it still makes sense to speak of the members of the smaller set as being weakly obligation-making. We can say that M1 is a strongly obligation-making motive when a statement of this form is true.

There is a motive M1 such that for every X which is not objectively obligatory, 'X from M1' is weakly obligation-making.

If there is a strongly obligation-making motive this also could come about in two ways. The first way is if there are objectively obligatory actions, and this motive makes all the other actions it favors obligatory. The idea of testing maxims itself does not determine whether there are objectively obligatory actions, but it leaves room for this possibility. The second way there could be strongly obligation-making motives is if there are no objectively obligatory actions. In that case, the motive is weakly obligation-making with respect to every action it favors. If there is such a motive then again mentioning the action it favors in a maxim is in a sense otiose. Whatever action it favors (given the failure of maxims favoring its omission) is obligatory. If there are either weakly or strongly obligation-making motives then PRC is false and MM is true. Given the three

Kantian arguments mentioned in Chapter 1 this would be a surprising result. But the idea of testing maxims itself allows for it.

The possibility of obligation-making motives is the obvious way that PRC would be falsified. This possibility was the focus of Prichard and Ross, and they supported PRC precisely because the idea of an obligation-making motive seems to violate the idea that we always can do what we are obligated to do. (That is, it violates the dictum ‘ought’ implies ‘can’.) However, wrong-making motives could also falsify PRC, and thereby undermine the possibility of objective obligations. Suppose that there is a strongly wrong-making motive M1. This means that any action that M1 favors is wrong. Hence, if M1 favors performing an action X1 that would otherwise be obligatory then X1 from M1 is wrong, not obligatory. For example, if malice is a strongly wrong-making motive then telling the truth or keeping a promise from malice is wrong. With such motives PRC is falsified not because there is one motive that makes X1 obligatory (or a few motives that make it obligatory). Instead, if X1 is performed from most motives it is obligatory, but if it is performed from M1 then it is not obligatory, but wrong. The specification of the obligatory action thus has the rider ‘unless it is performed from M1’. Even if a motive is weakly wrong-making it could make an otherwise obligatory action wrong.

Analogous remarks hold for obligation-making motives. If there are strongly obligation-making motives then no action is objectively wrong. If, say, sympathy is strongly obligation-making then lying or breaking a promise from sympathy would be obligatory, not wrong. I will not be discussing obligation-making motives until Chapter 8. The idea of an obligation-making motive is, in any case, rarely-encountered. The idea of wrong-making motives, as we have seen, is more familiar and even plausible.

I will summarize what we have found. Once we understand the concepts of a maxim and the testing of it for permissibility we see how Kantian theory could in principle yield a largely objective conception of deontic status. But the Kantian program of maxim testing leaves room for wrong-making motives (both weak and strong), and even—despite the Kantian arguments favoring PRC—obligation-making motives (weak and strong). So the theory need not yield a completely objective conception of deontic status.

## II

### The Practical Contradiction Test of Maxims

Nothing has been said yet about how the testing of maxims could show the deontic relevance of motives. We need to look at its inner workings of the test.

We have seen two statements that Kant offers of the Universal Law formulation of the Categorical Imperative. He goes on to illustrate it with four examples. Kantian scholars have used these materials, and others, to interpret the formulation and thereby articulate the nature of the moral reasoning it embodies. We will now look at the best interpretation known to me of how this test of maxims is supposed to work. This is Korsgaard's interpretation of the Universal Law formulation. She calls it the 'Practical Contradiction Interpretation'.<sup>21</sup> It has not convinced every scholar, but I think that for a number of reasons it is well worth our attention. For one thing, it has some independent plausibility as a form of moral reasoning. Second, Korsgaard explicitly insists on the relevance of the agent's reasons. Third, she gives an interesting interpretation of Kant's

idea that when a maxim fails the test a sort of contradiction results. Finally, Korsgaard's interpretation has a certain relation to its main competitor, the Logical Contradiction Interpretation. I will discuss how the relevant result in Korsgaard's interpretation applies a fortiori to this other interpretation. In order to avoid cluttering the text, I relegate that discussion to the notes.

So, to recapitulate the focus of this chapter: we will examine one interpretation (the Practical Contradiction Interpretation) of one formulation (the Universal Law Formulation) of Kant's conception of the supreme principle of morality (the Categorical Imperative).

Korsgaard argues that the test of maxims by the Formula of Universal Law (FUL) should be understood as asking whether what she calls a 'practical contradiction' occurs in a certain sort of imaginary situation. A 'practical contradiction', as Korsgaard explains it, is "a thwarted purpose".<sup>22</sup> It is not a theoretical contradiction of the form  $p$  and  $\neg p$ . It is a situation in which I cannot achieve my purpose in acting. The test, as she interprets it, works as follows. An agent's maxim will state, among other things, what her purpose is in acting, where this purpose is the agent's ultimate normative reason for acting.<sup>23</sup> The test will determine if this purpose could be achieved if she acted in a world where everyone tried to achieve it by means of the action she contemplates performing. I will call the imaginary situation in which the agent is acting the 'universalization situation'. If she could not achieve her purpose there, then acting on that maxim involves a practical contradiction. This shows that the act is impermissible.<sup>24</sup> Of course, the other side of the coin is that if her maxim passes the test then she is permitted to act on it. Note that the test does not find acting on a maxim to be wrong if there is a (practical) contradiction in

the maxim itself. Nor does it find acting on the maxim to be wrong if there is a (practical) contradiction in the description of the universalization situation itself. The practical contradiction occurs if we cannot conjoin a description of the universalization situation with the supposition that the agent succeeds in achieving her purpose there.

Korsgaard writes that “what the test shows to be forbidden are just those actions whose efficacy in achieving their purposes depends on their being exceptional.... Intuitively speaking, the test reveals unfairness, deception, and cheating.”<sup>25</sup> There does seem to be something morally problematic in my action if I could only succeed in performing it when other agents are not using my means of reaching my goal. The Practical Contradiction Interpretation of the FUL test therefore seems to be articulating morally significant features of rational action. It is thus at least a plausible form of moral reasoning. Herman has said that a test of moral permissibility should make clear why actions are wrong, when they are.<sup>26</sup> I think Korsgaard’s interpretation of FUL can make a claim to do this. We do well, therefore, to examine it carefully.

Kantians tend to picture two sorts of social process operating in the universalization situation. I will call the first a ‘victim protection process’, or Veep for short. A Veep involves intended victims of an action taking steps to thwart their victimizers. This is the process that is imagined in the well-known case of the ‘lying promise’: an agent is considering whether to state to a potential lender that she intends to repay the loan when she does not.<sup>27</sup> Lenders in the universalization situation will be the victims of this commonly-occurring act. Kant’s point is that they are rational and will take steps to protect their money, steps that will ensure that borrowers seeking to defraud them in this way fail to do so. Lenders who know that borrowers always lie about their

intentions will do things like dispense with asking for such statements, given their worthlessness. Lying in this fashion to obtain money in the universalization situation will therefore not succeed in obtaining money. The maxim of the lying promise thus fails the test of moral permissibility. Note that the issue in the test is whether the intended victims in the universalization situation will guarantee that the agents attempting to defraud them fail to achieve their purpose. The issue is not whether these agents will guarantee their own success in that situation. If the intended victims in the universalization situation guarantee that the agents fail then it is wrong for the deliberating agent to act on that maxim in the actual world.

I will call the second social process a reversed role process (or Reep). Korsgaard pictures this sort of process when she discusses another example. It involves an agent competing for a job who has the following maxim: ‘I will kill the best-qualified candidate for a job in order to get it myself’. She says of an agent with this maxim, “universalization makes you the victim.”<sup>28</sup> That is, in the relevant universalization situation the agent herself would be victimized by someone else’s operating on her maxim.<sup>29</sup> Reeps, of course, make the universalization test seem like the Golden Rule. One difference is that the Golden Rule is often phrased as seeking to know if the agent would resent or prefer not to be treated in the specified way. Korsgaard is asking whether the agent could succeed in reaching her goal if the behavior were directed at her.

Kantians sometimes picture a Reep thwarting the agent, and sometimes a Veep doing so. They also tend to picture only one of these processes operating in the universalization situation for a given maxim. There is in fact a problem with picturing both processes occurring. The supposition that both occur—to such an extent that the test

yields the verdict of impermissibility—leads to a (logical) contradiction. Consider a maxim that somehow harms others. A fully effective and disseminated Veep means that no instance of the contemplated act will succeed in the universalization situation: potential victims guarantee its failure. On the other hand, if the contemplated act is practiced in the universalization situation, the idea of a Reep pictures the agent herself being victimized—that is, thwarted—by the action of someone else, who succeeds when she acts on the agent’s maxim. However, if there is an effective Veep then the agent could use it just as well as anybody else, and hence she would not be a victim. Or, looking at it the other way, if the agent is a victim then there isn’t a fully effective Veep available. There is nothing contradictory in supposing that agents engaged in harming others take measures to avoid being harmed themselves: members of the armed forces do this. Likewise, there is nothing contradictory in supposing that burglars put locks on their own doors. What is contradictory is supposing that someone utilizes a method that guarantees she will not be harmed by a certain behavior and that she is harmed by that behavior. This tension between the two social processes does not invalidate the FUL test of maxims: a maxim could fail it. The important point is that if a maxim does fail it, this will be guaranteed by a Veep or a Reep, but not both.

Notice that with Reeps we are to picture other agents having the same motive (and the same maxim) as the deliberating agent. So if she is acting from a conditionalized maxim of revenge, say, then they are too. This is not the case with Veeps. If S has a malicious motive her victims might be motivated by self-interest when they try to protect themselves. The Formula of Universal Law test itself does not stipulate the maxims guiding these Veeps.<sup>30</sup>

### III

#### Reconceiving Motives

It would be convenient if we could proceed directly to contrast the consequentialist and Kantian treatments of the deontic relevance of motives. But there is an important preliminary issue that we should examine, namely, the conception of a maxim that the Practical Contradiction Interpretation commits us to. It actually requires maxims to have a certain form if Reeps and Veeps are to occur. We will focus on hostile motives, since some of these are plausible candidates for being wrong-making.

**Maxims that Generate Veeps.** In the permissibility test everyone in the universalization situation accepts the conditional maxim that the agent is considering, and acts on it when appropriate. The emphasized phrase brings out an important point about acting on a maxim. On a given occasion S may play Scrabble in order to enjoy herself. But she may not seek enjoyment when she is at work, or when the bills have to be paid. So, she may only be prepared to act for this reason in certain circumstances. This is more support for the idea that we should think of maxims as having the form ‘Xing from M in C’ (or, ‘I will X in C in order to achieve M’).

However, this form is not an accurate characterization of every action from a motive. Consider the motive of anger. Anger takes harm or insult to another person to be the end of action. In the standard case S’s acting in anger will involve her belief in, say, the desirability of insulting T as an end, and her intending to insult T as an end. But it need not involve implementing a plan to insult T if a certain condition obtains. It is not true that whenever someone does X from M in C she is implementing a conditionalized

maxim of the form, 'I will X from M in C'. Acting in anger is one such example. Kissing one's mistress from love is another.

How might anger be incorporated into a conditionalized maxim? One way to picture the psychology of an agent's formulating such a maxim is to suppose that S postpones her hostile act until a less risky situation materializes. She might thus accept a maxim of the form, 'I will insult T when she is unable to retaliate' (or, 'it will be desirable to insult T as an end when she is unable to retaliate'). Here prudence is modifying S's practical reasoning. We can also imagine other normative considerations underlying such conditions. S might adopt a maxim of insulting T when it will be particularly embarrassing to T. We actually have some hesitation in saying that if S later implements the maxim 'I will insult T when she is unable to retaliate' that she will then be acting in anger. However, this might not be thought to be a serious problem, since the same feature of the action is being characterized as desirable.

This conditional maxim is still not suited for the FUL test, though. The maxim 'I will insult T when she is unable to retaliate' does not have sufficient generality to activate widespread Veeps: if everyone acts on it when appropriate, only T will be victimized! Furthermore, adopting this maxim represents a commitment to act on only one occasion; the phrase 'when T is unable to retaliate' picks out that occasion. If potential victims are to develop successful thwarting strategies there must be recurrent patterns of behavior that they can accurately assess and then counteract.

One way to generate a recurrent pattern is to suppose that S is considering whether to affirm a maxim of revenge, like 'whenever I need not fear retaliation it is desirable for me to insult any person who insults me'.<sup>31</sup> Accepting this doubly quantified

maxim commits the agent to repeatedly acting in a certain way.<sup>32</sup> Maxims are now not merely conditionalized; they also contain universal quantification with regard to the condition and the action. Notice that if S repeatedly acts on such a maxim then she accepts that the same features of these actions are desirable as she does when she adopts the maxim, ‘it will be desirable [on one occasion] to insult T, who insulted me, when she is unable to retaliate’. In fact, if S affirms the generalized maxim we can picture her as engaged in constitutive practical reasoning when she chooses to insult T. The desirability of insulting T on this occasion is derivative.<sup>33</sup>

There might be ways to generate effective Veeps in the universalization situation that do not involve conceiving of rational agents as always implementing doubly quantified maxims. Nonetheless, it does seem that this is the simplest way to flesh out the idea that fully effective thwarting strategies will sometimes come about there.

**Maxims that Generate Reeps.** Reeps create further requirements concerning the form of maxims. If Reeps are to occur we must only consider motives that are formulated without referring to specific individuals or groups. Consider racist motives. Suppose a racist pushes someone aside who will not yield to her on a sidewalk. Let us grant that she is implementing a doubly generalized maxim. We can suppose she acts on this one:

RM1: Whenever doing so is not very costly or risky I will push black people aside if they do not give way to me.

If everyone in the universalization situation acts on RM1 when appropriate we find that absurd results ensue. Making this supposition seems to mean that blacks will act on the maxim and thereby victimize members of their own race. But the critical problem is this: if everyone acts on RM1 when appropriate then no white person will be victimized. Of

course, even if we imagine a universalization situation in which agents are thus harming blacks there will be Veeps that blacks undertake. And we can imagine they will somehow retaliate if the costs of doing so are not high. But neither of these social processes are Reeps: they are not reversed role processes in which blacks operate on the same maxim as whites and thereby victimize them.

It is instructive to devise a racist maxim that would generate suitable reversed role processes. Consider:

RM2: My acting to keep or to help establish my race in a superior social position is desirable as an end whenever doing so is not very costly or risky.

The phrase ‘my race’ is crucial here. In order to generate Reeps it must be interpreted to mean ‘whatever race I belong to’, rather than a specific race. That is, the phrase must be taken de dicto rather than de re. Otherwise, an agent might believe that it is desirable to establish, say, the Aryan race in a superior position. Many Nazis believed in the desirability of Aryan social superiority, but they probably only accepted the desirability of one named race’s domination, not the domination of whatever race they belonged to. That is, they accepted a maxim like RM2 in a de re sense. To adopt RM2 (in either sense) is to accept the desirability of contributing to a long-term strategy of racial domination. But the test of maxims only generates Reeps if we take RM2 to have a de dicto sense. If everyone in the universalization situation acts on RM2 in its de dicto sense there will be a contest for racial domination, and members of the Aryan race will be subject to Reeps.<sup>34</sup>

Notice that a maxim like RM2 (in either reading) is incomplete: it does not contain any reference to a specific action. In the next section I will utilize it to construct the maxim of an action.

In summary: the FUL test of maxims, as Korsgaard interprets it, does not allow us to determine directly the deontic relevance of all the motives that people actually act on. To generate Veeps we need maxims that are universally quantified in two respects. To generate Reeps we sometimes need to reinterpret the meaning of agents' beliefs about what is desirable in acting. If a racist's maxim is to guarantee that activities arise that threaten the agent herself then what she believes is desirable in her actions must be interpreted as having a de dicto meaning. Here the FUL test may require us to test a maxim that no agent has ever acted on. Both sorts of changes turn ultimate premises in practical reasoning into derived premises.

#### IV

#### Another Test Case

In order to present as close a comparison with the consequentialist treatment of motives as possible, I propose to explore the Kantian treatment of a version of the example discussed in the last chapter. The comparison actually cannot be very close. Kantian maxims sometimes require reformulating the motives that human agents actually act on. Furthermore, the testing of these maxims for our purposes would need to have a more elaborate structure than the example presented before. We are looking to see if Kantianism can affirm MM, and we have focused on the idea that a motive like racism is wrong-making. We are assuming that a verdict that racism is at least weakly wrong-making with respect to an act like refusing to shake someone's hand would tend to show that Kantianism can support MM in a plausible way. This suggests that we simply run the

FUL test on two maxims, one where a refusal to shake hands stems from self-interest, and one where it stems from a motive like racism. We would expect the first maxim to pass the test, and the second to fail it. But such a result would not prove that racism is even a weakly wrong-making motive with respect to such a refusal. After all, it is possible that many maxims, incorporating motives with widely different desirability characterizations, fail the test. To show that the maxim of a refusal from a racist motive is at least weakly wrong-making, we would need to test every possible maxim favoring the refusal. If we found that one or only a few motives, similar to racism in their normative content, played a role in maxims that failed the test we could conclude that they are at least weakly wrong-making, and possibly strongly wrong-making, with respect to the refusal. But in fact we can limit ourselves to running the test just once. This is because if the racist maxim passes the test then we can be sure it is not even weakly wrong-making with respect to the refusal.

Given what we found in the last section I propose to consider a case where S implements RM2 by the specific action of refusing to shake the hand of T, who belongs to another race. If S does this then she believes that it is desirable to contribute to a strategy of racial domination whenever a certain condition exists. RM2 should also be interpreted so that 'my race' has a de dicto meaning. Since S's refusing to shake T's hand implements a strategy of domination it has derivative desirability. The maxim of S's action then is this:

RM3: In virtue of its implementing my policy of acting to keep or to help establish my race in a superior social position whenever doing so is not very costly or risky, it is desirable to refuse to shake T's hand.

RM3 will actually pass the FUL test, not fail it. We can approach this conclusion from two directions, concerning the operation of Reeps and Veeps. Let us begin with Reeps. Given the stipulated meaning of the maxim we know that S will sometimes be subject to the sort of insult that she is carrying out. But even if this occurs often—and it might not, given the condition concerning risk—it will not make it impossible for her to do it on future occasions. No matter how often S is insulted she can still insult someone else.

It follows almost immediately that there is no fully effective thwarting strategy that Veeps will bring about. It is certainly hard to see how an agent could guarantee that she will not be insulted. There is admittedly a special issue about refusing to shake someone's hand: S can only do this when T is trusting enough to offer her hand to S in the first place. If we picture widespread racial animosity this sort of trust would not exist. But certainly all sorts of racial insults could occur in any case. This does not quite settle whether it is possible to guarantee that agents acting on RM3 will fail. The motive of RM3 does not take insult to be desirable as an end. The relevant insult is desirable in virtue of furthering a strategy of racial domination. Here again, though, it is hard to see how potential victims could guarantee that other agents will not be able to further this goal even slightly.<sup>35</sup>

## V

### Some Conclusions

Our most important conclusion is this: the best interpretation of the Kantian Formula of Universal Law fails to confirm MM in a plausible case. The program of testing maxims for permissibility could in principle show that MM is true. But when we formulate a maxim that would roughly parallel the action from a racist motive that we examined in the last chapter we find that the FUL test yields the verdict that the action is permissible, not wrong. Admittedly, we have not examined every motive that might be thought to be wrong-making. And we have not explored at all whether the Kantianism of the FUL supports Prichard and Ross's Claim (PRC). But I think we have enough material before us to say, tentatively, that consequentialism has a better way of treating the deontic significance of motives than does the Kantianism of FUL. We will need to look at some other features of Kantianism in the next chapter before we can reach a more complete assessment of the whole theory.

Before we move on, though, I think there are some further lessons that we can draw from our examination of the FUL. First, we should reconsider the contrast noted at the beginning of this chapter between consequentialism and deontology, and the significance that they give to the intrinsic features and effects of actions. The Kantianism of the FUL responds to the effects of actions. The Practical Contradiction Interpretation of the FUL test is seeking to know if one sort of effect, *viz.*, the frustration of the agent's purpose, will occur if the agent acts in the universalization situation. And we saw that Korsgaard is interested in whether two sorts of social processes (Reeps and Veeps) in the

universalization situation bring about this effect.<sup>36</sup> This test could show that motives are deontically relevant, but if it did so that would be in virtue of their own effects. This means that the Kantianism of the FUL is actually similar in a fundamental way to extrinsic consequentialism.

We can use utilitarianism as an example of extrinsic consequentialism to make this point clearer. The differences between the theories rest on two points. 1) The situations differ in which we examine the effects of a motive (that is acted on). Kant's test considers the effects of motives in the universalization situation, while utilitarianism considers the effects of motives in the actual world and in the alternative worlds open to the agent. 2) The sorts of effect that are relevant are different. Kant's test (as Korsgaard interprets it) makes the frustration of the agent's purpose in the universalization situation sufficient for impermissibility. Utilitarianism makes the production of less than the most happiness sufficient for impermissibility.

To say that the FUL test is responding to certain effects of motives is not to say it makes the deontic status of a particular act hinge on the occurrence of those effects in the actual world. The test says that acting on a given maxim is wrong even if doing so does not frustrate the purposes of the agent considering acting on it. S's lying promise is adjudged wrong even if she succeeds in receiving a loan. But the point stands that the test's operation in the universalization situation responds to the effects of motives there.<sup>37</sup>

If the FUL test of maxims finds motives to be deontically relevant because of their effects, then this part of Kantianism and extrinsic consequentialism both contrast with intrinsic consequentialism. Intrinsic consequentialism does not deny that motives are

deontically relevant because of their effects. But it does deny that they are only relevant because of their effects.

We will learn more by examining the social processes that bring about the relevant effects in the universalization situation. This helps us to see if the FUL test of maxims is after all a morally insightful way of testing an agent's practical reasoning. I granted above that there is some plausibility to this claim. But the conclusion we reached about the racist maxim invites us to reconsider it.

I think there is a sense in which Veeps presuppose the truth of consequentialism (if not utilitarianism). Consider their operation in the case of the lying promise. Kantians emphasize how rational lenders will take steps to protect themselves from such behavior. But why would they do this if they did not find something bad in losing their capital? When people fear that they will be subjected to behavior that pains or injures them they will usually make efforts to prevent this, if they can. But if, for some reason, they do not succeed in preventing the behavior that does not show that it is permissible. It may only show that the victims have no fully effective preventative measure. It is a contingent fact—when it is a fact at all—that rational agents can completely prevent the morally undesirable effects of others' behavior.

Reeps raise a different issue. In the Practical Contradiction Interpretation of the FUL it is the frustration of the agent's purpose that is the morally relevant effect. Frustration of the agent's purpose is not the same result as the one that utilitarianism finds relevant, namely, pain. But it is hard to see why frustration (that is, thwarting) is in fact the critical issue. In the case of the refusal to shake hands we found that the agent could still succeed in performing such acts when she is subjected to them herself. But if

being subjected to them is painful (or humiliating) it is plausible to think that this is the basic reason that these acts are wrong.

These reflections support our tentative conclusion that consequentialism has a better way of understanding the deontic significance of motives. It not only supports MM in a plausible case; consideration of the moral reasoning that constitutes this support suggests that it is more insightful than the reasoning embodied in the Formula of Universal Law.

---

<sup>1</sup> RG 7. Prichard, “Does Moral Philosophy” *op. cit.*, pp. 1-17, at 6-7; Prichard, “Moral Obligation,” *op. cit.*, pp. 87-163, at 129-35. See also C. D. Broad, “Analysis of Some Ethical Concepts,” in *Broad’s Critical Essays*, *op. cit.*, pp. 78-80; Broad, “Some of the Main Problems of Ethics,” *Ibid.*, pp. 223-46, at 239-40 (first published, 1946). Cp. William Frankena, *Thinking About Morality* (Ann Arbor: University of Michigan Press, 1980), pp. 48-49.

<sup>2</sup> PMJ 91.

<sup>3</sup> Kant speaks of deriving “duties” from the CI. G 89-91 (421-3). Cp. G 96-98 (429-30). And his various formulations of the CI are often phrased as imperatives requiring us to ‘act’ in certain ways—which is most readily understood as employing the deontic concept of obligation or requirement. G 88 (421, top); G 89 (421, bottom); G 96 (429). Cp. G 101 (433): each rational “should treat himself and all others...” The Formula of Autonomy (G 98 (431)), however, is not phrased as an imperative, nor is the Formula of the Kingdom of Ends (G 100 (434)). Herman argues in one essay that the various formulations of the CI cumulatively develop a theory of value. But she allows that the CI is also a test of moral permissibility. PMJ 208-40, esp. 215. Onora O’Neill claims that the CI tests basically establish the moral worth of ‘maxims’, as she interprets them, not the permissibility of actions. I believe that this is an unusual interpretation. See her *Constructions of Reason* (Cambridge: Cambridge University Press, 1989), ch. 7, pp. 126-44, esp. 130. Cp. p. 86. See also Scanlon, *Moral Dimensions*, *op. cit.*, ch.3, sec. 2.

<sup>4</sup> Richard Dean, *The Value of Humanity in Kant’s Moral Theory* (Oxford: Oxford University Press, 2006), p. 55 n. (Hereafter abbreviated as VHK.)

<sup>5</sup> Allen Wood, *Kant’s Ethical Thought* (New York: Cambridge University Press, 1999), pp. pp. xx-xxi. (Hereafter abbreviated as KET)

<sup>6</sup> See G 103 (437). For the denial of equivalence see KET, chs. 3-5, esp. pp. 97, 139-41; and Wood’s recent lectures available at available at the Oxford Philosophy Faculty website: <http://www.philosophy.ox.ac.uk/misc/berlin/> Cf. Lecture 2, pp. 44-7. Andrews Reath, *Agency and Autonomy in Kant’s Moral Theory* (Oxford: Oxford University Press, 2006), devotes some effort to responding to Wood on this point. See chs. 5, and 7, *passim*. A brief series of points is presented on pp. 160-1 n. 7. For other important discussions of the relations of the formulations see PMJ, ch. 10; CKE 100-101; CKE ch.

---

5; O'Neill, Constructions, *op. cit.*, ch. 7; Thomas Hill, Dignity and Practical Reason in Kant's Ethical Theory (Ithaca, New York: Cornell University Press, 1992), pp. 48-50; 52. (Hereafter abbreviated DPR) Hill's interpretation of the second formulation finds that it does not answer many important questions and is not equivalent to the others. Herman, Korsgaard, O'Neill and Hill differ quite markedly on this crucial matter.

<sup>7</sup> Kant distinguishes between narrow (or strict) duties and wide (or broad) duties, and he states that the derivation of one type will differ from that of another. G 91 (423-4). This means that an interpretation of a formulation should be able to show some difference in how these two types of duty are grounded. Korsgaard considers it to be a problem for two rival interpretations of the Universal Law formulation that they cannot easily do this.

CKE 95-7. (Korsgaard explains the relation of the distinction of strict and narrow duties in Kant to the better-known distinction of perfect and imperfect duties at CKE 20-21.)

<sup>8</sup> G 70 (402). Emphasis deleted. Cp. G 88 (421, top).

<sup>9</sup> G 89 (421, bottom). Emphasis deleted. Commentators have noted the slight difference between the two statements, and given them the titles. They are taken to be variants of the first formulation of the CI.

<sup>10</sup> G 89 (422). Wood conveniently gives a list of six maxims found in Kant's own writings, along with some discussed by later writers at KET, p.xxii. It is noteworthy that few of these maxims are explicitly of the form discussed below, 'Xing in C from M'. But cp. Korsgaard's criticism of another sample maxim at CKE 82.

<sup>11</sup> This is clear at G 70-1 (402-3); G 89-91 (421-3). Cp. CKE 39, n. 25.

<sup>12</sup> Two older discussions of maxims are Barbara Herman, Morality as Rationality (New York: Garland Publishing, 1990), ch. 2; Henry Allison, Kant's Theory of Freedom (New York: Cambridge University Press, 1990), pp. 86-94. See also CKE, 13; 57; 82; 92; 104, n. 20; Christine Korsgaard, The Sources of Normativity, (New York: Cambridge University Press, 1996), pp. 242-3; PMJ 13, n. 19; 51; 64, n. 29; 75; 103; 143f; 187 n. 3; 217-24; Reath, *op. cit.*, pp. 19; 43; 45; 85-6 n. 9; 89-90 n. 28; 133.

<sup>13</sup> CKE 13. Velleman quotes this passage in support of his claim that "maxims...state the connection between reasons and actions." Self, op. cit., p. 124.

<sup>14</sup> PMJ 13, n. 19.

<sup>15</sup> PMJ 95-6.

<sup>16</sup> Nell, Acting, op. cit., p. 75. Nell there states that Kant nowhere spells this procedure out, but that it is implicit in what he writes, given deontic logic.

<sup>17</sup> "On a Supposed Right to Lie from Philanthropy," in Immanuel Kant, Practical Philosophy, trans. and ed. Mary Gregor (Cambridge: Cambridge University Press, 1996), pp. 611-15.

<sup>18</sup> CKE, ch. 5; PMJ 149; KET 349-50, n. 14; DPR ch. 4; VHK 218-20.

<sup>19</sup> In an interesting discussion Herman suggested that the procedure of maxim testing be understood as first of all establishing a presumption that certain actions (like lying) performed from self-interest are wrong. PMJ 147-51. This establishes a moral presumption against performing them. She believes that this presumption can be rebutted.<sup>19</sup> PMJ 149. That is, Herman believes that the Categorical Imperative does not lead to rigorism. But she does not state that many maxims of self-interest should actually pass the test.

---

<sup>20</sup> Mark Timmons reaches this conclusion after examining MetM. Timmons, “Motive and Rightness,” *op. cit.*, pp. 255-88.

<sup>21</sup> CKE 77-105, esp. 92f. The test determines moral permissibility: CKE 152.

<sup>22</sup> CKE 96. This notion of contradiction may seem strained. It does occur elsewhere in the literature. See Anscombe, *Intention*, *op. cit.*, p. 55.

<sup>23</sup> I am equating ‘purpose’ and ‘reason’, though Korsgaard does not herself explicitly do this. But we saw (at notes 12 and 13 above) that she requires the maxim to contain a statement of the agent’s reason.

<sup>24</sup> This version of the test establishes that an action is a duty of narrow obligation.

<sup>25</sup> CKE 92. Cp. G 91-2 (424).

<sup>26</sup> PMJ 115-16.

<sup>27</sup> G 70-71 (402-3); G 89-90 (422); G 97 (429-30).

<sup>28</sup> CKE 98.

<sup>29</sup> Korsgaard adds that an agent who kills with this maxim not only has the purpose of getting the job, but also of being secure in its enjoyment over time. It is this latter part of the agent’s purpose that is thwarted by universalization. CKE 99.

<sup>30</sup> The main competitor to Korsgaard’s interpretation of the Universal Law formulation of the CI is what she calls the Logical Contradiction Interpretation. CKE 81-87. It says that in some cases a logical contradiction (i.e., a conjunction of the form  $p$  and  $\neg p$ ) exists if the agent acts on her maxim in the universalization situation. So, in the case of the ‘lying promise’ universalization will lead to there not being promises to repay at all. The picture here is of creditors becoming so skeptical that the practice of promising disappears altogether (at least in the case of borrowing). Presumably this is because promising is a way of creating an obligation to do something by means of intentionally producing an expectation in another person that you will do it. But if S knows that the person who hears her speaking will not believe that she is going to carry through, then she cannot intend to create the expectation that she will do so. So S cannot promise to repay. There is then a logical contradiction in supposing that there are no promises to repay and that S promises to repay. Once again, the interpretation is not looking for a logical contradiction in the maxim itself, nor is it looking for one in the universalization situation. If there is a contradiction it occurs in conjoining a description of the universalization situation and the assertion that the agent acts on her maxim.

I think it is fair to say that this interpretation is picturing the same Veeps going on in the universalization situation as is the Practical Contradiction Interpretation. One might contrast the two interpretations by saying that in the Logical Contradiction Interpretation Veeps ‘go farther’ so that, for example, it is not merely the case that a lying promise cannot succeed, but that it does not exist at all. Reeps are a different matter. If Reeps are occurring then, by hypothesis, the agent’s behavior has not disappeared: the agent herself is being subjected to it. The crucial point for our purposes is this: if a maxim passes the Practical Contradiction Interpretation of the test, it will pass the Logical Contradiction Interpretation *a fortiori*. If S will not fail to achieve her goal by means of Xing in the universalization situation, then it is not the case that Xing does not exist in that situation.

<sup>31</sup> Cp. the simpler maxim of revenge in Kant, *Critique of Practical*, *op. cit.*, 17 (19).

<sup>32</sup> A maxim of suicide is an obvious exception. There is thus something odd in the language Kant uses to formulate a sample maxim about this. “From self-love I make it

---

my principle to shorten my life if its continuance threatens more evil than it promises pleasure.” G 89 (422), my emphasis.

<sup>33</sup> Mele notes that even a derivative desire to, say, avenge T’s insult, might characterize this act as desirable as an end. Mele, Motivation, op. cit., p. 132.

<sup>34</sup> This is obviously related to the role of “the veil of ignorance” in TJ 136-42.

<sup>35</sup> It follows that RM3 passes FUL under the Logical Contradiction Interpretation. The Veeps that would occur in the universalization situation would not make it the case that no refusals to shake hands take place there.

<sup>36</sup> These sorts of processes also occur in the Logical Contradiction Interpretation of the Universal Law test. Where it differs is in the further effect of them it is interested in, namely, the disappearance of a certain activity.

<sup>37</sup> Wood remarks that “(hypothetical) consequences” are important in applying the Universal Law test. KET 92. Cp. 88.