

Chapter 5

Treating Humanity as an End

The best-known statement of the second formulation of Kant's Categorical Imperative—usually called the Formula of Humanity—is this:

FH: "Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end."¹

On the face of it, the Formula of Humanity and the Universal Law formulation are different, since the former principle does not mention maxims², nor does it invite us to consider what happens if the agent acts in the universalization situation. FH is now often said to be the heart of Kantian moral theory, and to provide a more readily defensible or understandable principle for deciding what it is right to do.³ Once again, we cannot explore all of the issues surrounding the interpretation of the Formula of Humanity, but must limit ourselves to understanding how it bears on MM.

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.

As in the previous chapter, I will utilize what I take to be the main line of interpretation of FH in the body of the chapter.⁴ Richard Dean has recently presented a dissenting interpretation of it that I take up in the notes.⁵

In examining FH we need to keep some things in mind. First of all, we are to interpret it as the single fundamental principle of morality. This means that FH will establish the deontic status of all actions: which are obligatory, which wrong, and which merely permissible.⁶ The Formula of Universal Law is also offered to us in this spirit. Not only should FH yield deontic judgments: these should be generally congruent with what we think about these matters. For our purposes one aspect of this congruence is especially important. FH should yield a largely objective conception of deontic status. That is, FH should yield deontic judgments that make motives irrelevant in most cases. We saw how FUL could in principal confirm this by showing, for example, that a maxim of telling the truth from self-interest passes the test of permissibility. FH should be interpreted in such a way that it too entails that an agent can carry out such an obligation from various motives.⁷ It will take some work to develop such an interpretation. I do this in Sections I-III of this chapter.

On the other hand, if Kantianism is to support MM it must also say that sometimes the motive of an action is relevant deontically. And it should say this in cases that have some independent plausibility. We saw that the Formula of Universal Law failed to do this in our test case. In Sections IV and V I will show that FH also fails to meet these requirements. It fails to do so in a starkly contrasting way, though. FH makes malice a strongly wrong-making motive, which is also an unacceptable position.

I

Treating People as Means

In this section I explain the significance of some of the terms in FH and begin our search for a largely objective interpretation of it.

FH tells us, in part, not to treat ‘humanity’ simply as a means. Commentators on Kant generally agree that—curiously enough—the reference to ‘humanity’ in FH signifies rational nature.⁸ This makes sense, though, when we recall that Kant repeatedly claims that the fundamental principle of morality is binding on all rational beings.⁹ And, indeed, Kant later restates FH thus:

...rational beings all stand under the law that each of them should treat himself and all others never merely as a means, but always at the same time as an end in himself.¹⁰

Furthermore, the commentators generally (but not entirely) agree that the essential feature of rational nature for Kant is the capacity to set ends and judge the desirability of goals for one’s self.¹¹ Korsgaard explains Kant’s meaning:

...the distinctive feature of humanity, as such, is simply the capacity to take a rational interest in something: to decide, under the influence of reason, that something is desirable, that it is worthy of pursuit or realization...It is this capacity that the Formula of Humanity commands us never to treat as a mere means, but always as an end in itself.¹²

These glosses mean that FH requires us never to treat rational beings ‘simply as means’, and always to ‘treat them as ends’. We now must determine what these latter two expressions mean. We can begin with the expression ‘treating someone simply as a

means'. Since Kant incorporates this idea in his formulation of the fundamental principle of morality, we might conclude that this is the basic condition of moral permissibility. In that case, using people simply as a means is necessary and sufficient for wrongdoing, and not using people simply as a means is necessary and sufficient for permissibility.

However, under a natural interpretation this focus on the significance of means falls outside of our topic, since it does not concern an agent's motive. Motives are the ultimate premises in practical reasoning. They determine the agent's end in acting. The agent's means are derivative considerations normatively; in any formal presentation of practical reasoning they would occur after (or under) the statement of the agent's motive or motives. Kant may well have supposed that self-interest is always the motive of someone who treats another person simply as a means, but this need not be true. Kidnappers often treat their victims simply as a means, but they may have political or religious motives for doing so. This is not to say that an agent's means are never relevant deontically. It is only to say that if her means are deontically relevant this would not be in virtue of MM.

We can nevertheless learn some interesting lessons if we examine the issue of the deontic relevance of an agent's means. Consideration of the prohibition on using rational agents simply as means will help us to find an interpretation of FH that yields a largely objective conception of deontic status.

FH rules out using people simply (or merely) as a means. We can examine this idea by first considering the idea that using people as a means is the basis of deontic status. We can then ask about the significance of the modifier 'simply'.

Let us consider whether using people as means is necessary and sufficient for acting wrongly. This would mean that the fundamental principle of morality asserts two things:

A1. We are not permitted to perform actions that treat people as means;

B1. We are permitted to perform actions that do not treat people as means.

If these statements are utilizing the term ‘means’ in its ordinary signification it is clear that they are not true. As far as statement A1 goes: using people as means is not sufficient for wrongdoing, since we permissibly use people as means in many ways. We permissibly determine what time it is by having others tell us; we are permissibly transported by bus drivers and airline pilots; we are permissibly served food by waiters.

Statement B1 is also false: using people as means is not a necessary condition of wrongdoing. There are numerous counterexamples to this claim. I will mention four.

- 1) As Jonathan Bennett in effect noted, it is false to say that treating someone as a means is necessary for wrongdoing when harm to her is an unintended but foreseen effect.¹³ (Such effects are sometimes called ‘double’ effects, or ‘collateral damage’.) If S intends to set off a bomb in her backyard for fun, and knows that this will kill her neighbor T, this further effect may be an unintended byproduct of her activity. A fortiori S is not using T as a means. And yet this action is wrong.
- 2) Harm to a person can be produced as an end, not as a means. This is precisely what malice as a motive seeks to do. S may find it to be desirable as an end to

burn down T's house or to kill T. The fact that doing these things serves no further purpose of S does not mean that they are permissible.

- 3) A failure to help another person generally is not an example of using her as a means. If S is a Bad Samaritan and walks by T, who is suffering in the road, she does not use T as a means.¹⁴ But S's action is wrong. This point applies in the case of self-regarding acts as well. Kant believes that omitting to develop one's talents is wrong. But to omit to develop these talents is not to use one's self as a means.¹⁵
- 4) Failures to impose obligatory harms such as punishment are problematic. In Kant's celebrated example of the 'desert island' a governor fails to impose the mandatory death penalty on a guilty murderer.¹⁶ If we grant to Kant that this action is wrong, we do not find its wrongness to be due to the fact that the governor is using the prisoner as a means. The governor is not using the prisoner as a means. The point applies more generally. Even if we don't agree that in this case the governor acts wrongly, we do believe that sometimes it is wrong not to impose a punishment. And such inaction is not an example of using the criminal as a means.

In all of these cases the problem stems from insisting that a certain structure in S's practical reasoning is a necessary feature of wrongdoing. But—we want to say—S acts wrongly even if this structure is not present. S is not using T as a means when she blows up her bomb and kills T, but this does not entail that S acts permissibly.

To assert that using someone as a means is necessary and sufficient for wrongdoing does not make the agent's motive the only deontically relevant fact. But it makes another feature of the agent's practical reasoning the only deontically relevant fact. It thus makes deontic status completely subjective, that is, dependent only on this feature of the agent's practical reasoning. This is implausible.

II

Treating People Simply as Means

We can now consider how the word 'simply' changes our conclusions. Is it reasonable to think that the fundamental principle of morality rather asserts these two things?

- A2. We are not permitted to perform actions that treat people simply as means;
- B2. We are permitted to perform actions that do not treat people simply as means.

The change from B1 to B2 does nothing to help with the four counterexamples to B1. If S treats T simply as a means then she treats T as a means. So if S does not treat T as a means then she does not treat T simply as means. Hence, B2 also tells us that the actions in the four examples are permissible. These are again implausible answers.

A2 is not so easily disposed of. There is some plausibility in claiming that if S does treat T simply as a means then she acts wrongly. I suggested above that kidnapping seems to be an example of using someone simply as a means, and we might say the same about rape. But we need to examine A2 more carefully.

Derek Parfit has recently argued in effect that A2 is false. His most penetrating point is based on the following assertions. Parfit says that treating someone simply as a means is based on an attitude of complete indifference to a person's welfare and claims.¹⁷ In order for S to be treating T simply as a means we need to add that S is acting in such a way that she uses T as a means in virtue of this attitude. Parfit claims that acting in this way is not, in fact, always wrong.

Consider some gangster who...regards most other people as a mere means, and who would injure them whenever that would benefit him. When this man buys a cup of coffee, he treats the coffee seller just as he would treat a vending machine. He would steal from the coffee seller if that was worth the trouble, just as he would smash the machine. But, though this man treats the coffee seller merely as a means, what is wrong is only his attitude to this person. In buying his cup of coffee, he does not act wrongly.¹⁸

Parfit notes that, similarly, a thorough egoist could keep a promise from self-interest.

Parfit believes that the fact that an agent has such an attitude of indifference is brought out by some counterfactual statements about her behavior. He says of the gangster, "he would steal from the coffee seller if that was worth the trouble." Needless to say, the vast majority of people who act on occasion from self-interest do not have this attitude about any other person. If S gives T the correct change only from self-interest then (in the standard case) she believes that furthering her self-interest (or some element of it) is desirable as an end; she believes that giving T the correct change furthers her self-interest; and she is trying to further her self-interest by giving T the correct change. From these facts it does not follow that S would steal from T, let alone kill or maim her, if she thought that doing so was in her self-interest. But Parfit holds that even if S did have such a comprehensive attitude of indifference towards the well-being and claims of

T, she would not be acting wrongly if she manifested this attitude in the act of buying coffee from T (in the usual way).

I think Parfit has stated correctly what we mean by saying that someone treated another person simply as a means. And I think the contention that his gangster does not act wrongly in buying coffee is plausible. I therefore agree with the assertion that A2 is false.

It is surprising to find a case where treating someone simply as a means is permissible. But if we wish to explore this result further, we can ask whether Kant's theory really does entail that the gangster acts wrongly. Kant partly acknowledges the largely objective nature of deontic status via the distinction between acting in accordance with duty and acting from duty. He claims that a merchant carries out her obligation to act honestly when she returns the correct change from self-interest. It follows that a merchant who treats her customers as means of making money by returning the correct change carries out this obligation. In fact, to say that S is returning the correct change to T from self-interest is just to say that by giving T the correct change S is treating T as a means of furthering her own well-being. So Kant himself is committed to saying that S acts permissibly, and not impermissibly, in performing this action that treats T as a means. It is not clear why Kant should then assert that Parfit's gangster acts wrongly in treating the server simply as a means. It seems that Kant should say that the gangster acts in conformity with duty, that is, permissibly.

This discussion is not conclusive. Perhaps there is some further argument in Kantian theory that shows that Parfit's gangster acts wrongly. Or perhaps Kant has another, technical conception of what treating someone simply as a means entails. But the

gangster case is not the only difficulty that exists for the claim that A2 and B2 constitute the fundamental principle of morality. And we need to remind ourselves that the claim that A2 and B2 constitute the fundamental principle of morality makes deontic status completely subjective—even if the crucial subjective factor is not the agent’s motive.

III

Treating People as Ends

These reflections suggest that the central idea in FH is really the requirement always to treat humanity as an end.¹⁹ It seems like a drastic reinterpretation of FH to say this, given its explicit reference to treating people simply as a means. But I think there is some textual evidence that this is Kant’s view.²⁰ I will show that this interpretation also has the virtue of yielding a largely objective conception of deontic status. This does not rule out the possibility that some actions are wrong in virtue of the fact that they involve treating people simply as means. One way that this could turn out to be true is this: in the limited number of cases where treating people simply as means (as this is ordinarily understood) is in fact wrong we can see that such actions also fail to treat people as ends. We can then say that FH entails in these cases that treating people simply as means has derivative relevance. We do not need to look at these cases, though, since an agent’s means do not constitute her motives.²¹ I will mention in a moment another way that treating people simply as means could turn out to be relevant deontically, given the truth of FH.

Let us now turn to the all-important issue of what it is to treat rational beings as ends. In trying to forestall a natural misunderstanding of his thinking Kant here does introduce some special terminology. We think of an ‘end’ as a state that an agent seeks to produce by her action. Kant sometimes speaks of this as a “subjective end”.²² It suggests once again that FH enunciates a completely subjective conception of deontic status, by requiring us always to seek one end or act from one motive. Kant makes a crucial distinction: he says ‘end’ has a different meaning in FH.

Persons...are not merely subjective ends, whose existence as an object of our actions has a value for us: they are objective ends—that is, things whose existence is in itself an end...²³

...the end must here be conceived not as an end to be produced, but as a self-existent end.²⁴

Kant thus claims that rational agents are a second sort of ‘end’ that morality properly recognizes. He also makes a compressed argument that connects the two sorts of end. The basic idea is that if an agent can rationally judge and set (subjective) ends then she herself must be an objective end.²⁵ (Kant also speaks here of “an end in itself”.²⁶) We could put his argument in our terminology by saying that an agent who tries to bring about a state of affairs that she believes is desirable is herself an objective end. This is a striking and strange use of the term ‘end’.²⁷ The German term corresponding to both senses of ‘end’ in the quotations is ‘Zweck’. But as far as I can determine Kant’s use of ‘Zweck’ to refer to existing individual persons is idiosyncratic, and Paton’s translation carries this over into idiosyncratic English.²⁸ Indeed, the phrase ‘Zweck an sich (selbst)’—perhaps meant to echo ‘Ding an sich’—is a forceful way to capture what J. L. Mackie called a value that is “objectively prescriptive”.²⁹ We might also say it captures a

certain conception of objective normativity, that is, the idea that these ends provide all rational agents with normative reasons to treat them in certain ways.

Moral requirements can therefore be thought of as proper responses to the distinctive value that objective ends have. For a given agent S another agent T constitutes such an objective end. Indeed FH tells us that S is herself an objective end with regard to her own activity.

Once we realize that FH is telling us to treat rational agents as objective ends we can see that it is not necessarily telling us to pursue any subjective end or act from any specific motive. But we need guidance in understanding the nature of the value that objective ends have, and which actions respond properly to it. Kant is not as helpful here as we might wish, since he gives no general characterization of the value that objective ends have. After arguing that we have this value he immediately proceeds to discuss four examples of duties that follow from FH. Commentators who try to expound Kant's general conception of this value must therefore use a variety of materials to piece it together.³⁰ Two brief remarks that Kant makes later are suggestive, though. The first follows his introduction of the idea of a "self-existent end" just quoted and it explains its relevance thus:

It must ...be conceived only negatively—that is, as an end against which we should never act...³¹

Kant also states that objective ends function as "the supreme limiting condition" on all the subjective ends that S sets for herself.³²

It is natural to take Kant to be saying that such a limiting condition corresponds to the deontic division between the impermissible and the permissible.³³ And it is natural to

take the first quotation to mean that actions which damage or impair a person's rational capacities, or interfere with her autonomous rational activity are impermissible.³⁴

However, it is clear that this is not enough material to build a complete moral system.

The idea that the appropriate response to objective ends is simply not to harm them or interfere with their rational activity entails that we have no obligations to help or benefit them. Kant clearly rejects this idea. He writes that

...humanity could no doubt subsist if everybody contributed nothing to the happiness of others but at the same time refrained from deliberately impairing their happiness. This is, however, merely to agree negatively and not positively with humanity as an end in itself, unless every one endeavors also, so far as in him lies, to further the [subjective] ends of others.³⁵

These quotations suggest how FH could serve to establish obligations not to kill or maim rational beings,³⁶ not to destroy or impair their rational capacities; not to coerce them; and to give them assistance in achieving the ends they rationally set for themselves.³⁷ (So although objective ends are radically different from subjective ends, objective ends give us a normative reason to pursue the subjective ends of others.³⁸) In the Groundwork Kant himself uses FH to establish other obligations, for example, of truthfulness or non-deception, and obligations to refrain from theft, but his reasoning here is, in my opinion, less clear and convincing.³⁹

If we worked out a system of obligations and prohibitions based on this idea of rational beings as ends in Kant's technical sense we might then regard the phrase 'treating people simply as means' as also involving a technical sense. We could say that all actions that are impermissible in virtue of failing to treat people as ends in Kant's technical sense thereby treat them simply as means in a correlative technical sense. This

suggestion is another way in which the idea of treating people simply as means could be claimed to have deontic significance in FH. Kant does not, as far as I know, ever assert that he is using the phrase ‘treating people simply as means’ in a technical sense. But since this suggestion requires that we first develop the implications of treating people as ends in Kant’s technical sense, I will continue focusing on that latter idea.

Let us grant, then, that this idea of treating persons as ends in themselves can be used to generate a certain system of obligations and prohibitions. The concepts involved in these obligations and prohibitions—deception, coercion, killing, assistance, for example—will be suitably objective in our sense: one can do these things (and refrain from doing them) from many different motives. S can carry out her obligation to assist T in furthering her ends ‘from duty’, but she can also do this from self-interest, and from many other motives.⁴⁰ S can carry out her obligation not to kill T from many different motives; and she can violate this obligation from many different motives. To say this is to say that objectivity does not only apply in the category of the obligatory. There is also objectivity in the deontic category of wrongness—and, presumably, mere permissibility. Kant suggests that obligation is objective by means of his distinction between acting in accordance with duty and acting from duty. FH enables him to say that it is possible to act wrongly, but not ‘from wrongness’, as it were.

The interpretation of FH that I have sketched is more helpful in understanding certain kinds of wrongdoing than the prohibition on using someone simply as a means (in its ordinary sense). One kind of wrongdoing it helps with is ‘double effect’ cases. If S blows up her bomb while foreseeing the death of T, her doing this can now be seen as violating FH. Her action destroys a rational being. Another kind of wrongdoing it helps

with is maliciously motivated violence. If S believes that killing T is desirable as an end, and intentionally kills T, this can be seen as violating FH. Again, S's action destroys a rational being.

Given what we have established thus far it seems that FH generates a system of absolute obligations and prohibitions. That is, it seems that FH entails, for example, that all killings of rational beings are wrong, and that it is always obligatory to refrain from killing them. Kant himself did not believe this since, as we saw, he asserted that capital punishment of murderers is morally obligatory, not morally wrong. Although he believed that all lying is impermissible, we saw that most contemporary Kantians reject this claim. If FH is the fundamental principle of morality then any exceptions to a general obligation or prohibition that it establishes must also derive from FH. Two concepts that seem to be related to the central idea of the rational setting of ends are consent and forfeiture. It might then be said that the prohibition on maiming disappears if a person rationally consents to it, for example, by agreeing to undergo surgery. It might also be said—as Kant did say—that the prohibition of killing T disappears if T rationally chooses to kill someone else. This can be taken to mean that T forfeits her right to life (i.e., her right not to be killed) if she rationally chooses to kill U. There are other ways to work exceptions into the obligations and prohibitions that FH generates that might be said to be based on the concepts contained in FH itself. For example, Kant himself sometimes appeals to the notion of what a person 'could not agree to' (as distinct from what she does not agree to).⁴¹ Obviously we cannot here determine which concepts, arguably contained in FH, can be used to generate certain exceptions to the moral rules that seem to flow from it. The important point for our purposes is that all the considerations just mentioned are

suitably objective. If T's consent establishes that S's cutting her is permissible, and not wrong, then this permissibility-establishing fact has nothing to do with S's motives in acting. Hence, S may cut T (when she consents to be cut) from many different motives. The same point applies if the permissibility-establishing consideration is the modal fact that T could agree to S's action.

We have now examined a very rough sketch of an interpretation of the principle requiring us always to treat persons as ends. But I think it enables us to see how the notion of treating people as ends might yield a largely objective conception of deontic status.⁴² This is a reason for supposing that it really is the central idea in FH. We can now consider how this interpretation bears on MM. We need to ask if it leaves any room for motives that are deontically relevant.

IV

Treating People as Ends and Wrong-Making Motives

I believe the best way to address this question is to return to the idea of a wrong-making motive. We have seen that some of the clearest cases of deontically relevant motives fall into this category. Kant may have held a completely objective conception of obligation; that is, he may have accepted PRC.

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

This leaves room for wrong-making motives. In this section I first consider how acting from the motive of malice can be seen as a violation of the Formula of Humanity. I then

look at the sparse textual evidence we have in Kant's work on this question. It seems to support what I call 'intrinsic Kantianism'. This form of the theory takes malice to be a strongly wrong-making motive, and hence supports MM.

In Kant's late work The Metaphysics of Morals he in effect treats the deontic significance of certain motives. The latter part of this work, "The Doctrine of Virtue", deals with duties that are not appropriately subject to legal requirements. It covers topics like duties to one's self, beneficence and gratitude. Here Kant discusses three vices that involve hatred of other people (envy, ingratitude and malice), and three vices that involve disrespect of other people (arrogance, defamation and ridicule).⁴³ In the next chapter we will see that virtues and vices often consist in part in distinctive patterns of motivation. This means that these particular vices generate distinctive motives, and Kant discusses them in relation to FH.⁴⁴ We will see in a moment that his treatment of motives in this work has significant limitations for our purposes, but we can still make use of it. It is, incidentally, an interesting fact about the history of moral thought that Kant does not discuss motives that we speak of as, for example, racist or anti-Semitic, and that I have suggested we now think of as wrong-making.⁴⁵ I will largely use Kant's treatment of the traditionally-condemned motive of malice in this discussion.⁴⁶ He seems to think of this as something directed at a particular individual, and stemming from personal animus, rather than as something directed at a class of people.

Let us begin by considering ourselves how a Kantianism based on FH might support the idea that malice is a wrong-making motive. Drawing on the texts quoted above, we might pose the problem this way: can we see how acting from malice is 'acting against' a person, that is, against 'an objective end'? We can use the example of refusing

to shake hands from malice. Let us use the variant in which S's malice is clear to T, and this is painful to her. The answer to our question is clear. FH requires us to treat persons as ends in themselves and S is violating it: she is 'acting against' T by exhibiting her hatred of T.

The reasoning just presented has a certain similarity to extrinsic consequentialism. Extrinsic consequentialism asserts that motives can be relevant deontically because of their effects. The Kantian reasoning just sketched also focuses on the effects of S's motive. It is saying that seeking to cause pain to a rational being as an end, and doing so, is wrong. This is not to say that the moral system that FH generates is some sort of consequentialism. For one thing, Kantianism is paying attention to effects highlighted by FH, rather than by, say, utilitarianism. The deontically relevant facts have to do with the way that effects on an 'objective end' can be seen as 'acting against' her. For another thing, Kantianism does not determine what is wrong by minimizing overall the effects it determines to be bad. To act against an objective end is wrong, whatever other effects the action may have. Keeping all of this in mind, we can devise a name for the position just expounded: 'extrinsic Kantianism'.

There is another position to which FH could lead, though.⁴⁷ It has a certain similarity to intrinsic consequentialism, since it focuses entirely on the significance of the motive guiding a malicious agent's activity.⁴⁸ We saw that intrinsic consequentialism says that some motives are intrinsically bad, and this badness exists even if it is concealed from the person who is affected by the agent's activity. Let us call the analogous set of claims intrinsic Kantianism. Intrinsic Kantianism says that whenever a person acts maliciously she 'acts against' a person; hence, she acts wrongly. This is true if the target

of S's malice is unaware of her motive, or even of her action. (Out of malice S might, for example, buy a statue that she knows her rival T wants, and T might never learn of this.) Intrinsic Kantianism is not identical to intrinsic consequentialism. Intrinsic consequentialism does not assert that every action from malice is wrong, since the motive of an action is only one factor in its moral calculation. Intrinsic Kantianism, in contrast, asserts that every malicious act violates the fundamental principle of morality. Acting with a certain 'subjective end' is to violate a principle that requires honor and respect for objective ends.

Both forms of Kantianism just described seem to support the idea that malice is a wrong-making motive, and thereby support MM.

I developed these ideas without referring to Kant's own words. We can now examine them. The textual evidence in "The Doctrine of Virtue" is sparse. But I think it gives some support to the idea that Kant accepted intrinsic Kantianism.

It is striking that Kant—in spite of his retributivism—condemns revenge, which he takes to be a form of malice. (He calls it "the sweetest form of malice".⁴⁹) His position is that in an organized state the morally required response to a violation of a person's rights must come from a public authority. And he asserts that even these authorities may not inflict punishment "out of hatred".⁵⁰

We can look at some other passages. Consider this passage about the vice of contempt:

At times one cannot, it is true, help inwardly looking down on some in comparison with others...but the outward manifestation of this is, nevertheless, an offense.⁵¹

I think that the most natural way to interpret this passage is that manifesting contempt to another person is wrong. This suggests extrinsic Kantianism. But compare this:

Envy ...is a propensity to view the well-being of others with distress, even though it does not detract from one's own. When it breaks forth into action (to diminish their well-being) it is called envy proper; otherwise it is merely jealousy...⁵²

This might be thought to refer only to openly envious action. But just before this passage Kant says that envy is a vice of hatred which is “not open and violent, but secret and veiled.”⁵³ This suggests that when S acts enviously to diminish T's well-being T may not know S's motive, or even that S has acted at all. The implication is that S acts wrongly in any case. And we saw that Kant states: “no punishment, no matter from whom it comes, may be inflicted out of hatred.”⁵⁴ This suggests that all acts of punishment from hatred are wrong, even when the hatred is not clear to its target.

Finally, there is Kant's discussion of malice. In speaking of 'egotism' he says that it can lead people to be pleased by the sight of others' suffering, since this makes them feel their own good fortune more strongly. And if someone undergoes a reversal of fortune from scandal this can please others by reinforcing their sense of their own rectitude. Kant condemns this as self-conceit, since it is “only...their good fortune ... [to have] escaped temptations to public vice.”⁵⁵ He says this of the pleasure taken in others' misfortunes:

...to rejoice immediately in the existence of such enormities destroying what is best in the world as a whole, and so to wish for it to happen, is secretly to hate men; and this is the direct opposite of love for our neighbor, which is incumbent on us as a duty.⁵⁶

(The duty of love for our neighbor is what presumably Kant described as one of ‘practical’ love.)⁵⁷ This passage explicitly treats only the occurrence of pleasure taken in others’ misfortune. Such pleasure, not being an action, cannot be wrong. But the context encourages us to think that acting from such hatred is wrong. Malice, like envy, is “not open and violent, but secret and veiled”. This implies that even when the malicious motive of an action is not known to its target, the action is wrong. So the textual evidence—such as it is—points to intrinsic Kantianism. If this is correct then it suggests that Kant holds that malice is a strongly wrong-making motive.

There are two obvious gaps in this line of interpretation, though. First of all, in “the Doctrine of Virtue” Kant is not discussing a motive like malice in its full generality. His remarks there are limited to its manifestations in spheres where he thinks legal coercion is inappropriate. But obviously malice could lead an agent to perform such actions as killing or maiming. These are treated, if only by implication, in the first part of The Metaphysics of Morals.⁵⁸ Nevertheless, we have tentatively concluded that FH is capable of yielding a largely objective conception of deontic status, which means that the wrongness of these sorts of actions can also be established by appeal to that principle. If that is correct then Kant’s remarks in “The Doctrine of Virtue” should be seen as supplementing his treatment of actions that are objectively obligatory and wrong. He can then be taken to be saying that every malicious action which is not objectively wrong is wrong because of its motive. In our terminology that is to say that malice is weakly wrong-making with respect to every action that is not objectively wrong. And that means that malice is strongly wrong-making.

There is a second gap in the interpretation which concerns the issue of exceptions. FH could in theory allow for exceptions to any rule it established. So it is possible that FH establishes both that acting from malice is wrong, and that in certain circumstances it is not. If that were true then malice would not after all be strongly wrong-making, although it could still be weakly wrong-making. This possibility does exist, but it is hard to see how it would be realized. Kant himself seems concerned to place acting from certain motives up against the fundamental principle of morality, as it were. And one of his verdicts seems to be that malicious activity as such is opposed to the orientation towards rational beings that FH enjoins. How then might a further appeal to FH show that malicious actions are sometimes permissible? The most plausible suggestion, I think, is that some rational choices are such that the agents who make them forfeit the right not to be treated maliciously by others (or, perhaps, forfeit the right not to be treated maliciously by some others). This does not seem to be Kant's own view, though. He asserts that even criminals who are subject to severe punishments may not be punished "out of hatred". It is reasonable to think that if anyone has forfeited a right not to be treated maliciously it is these criminals, but Kant in effect denies this. A full discussion of this second point would require an understanding of all of the ways that FH could be used to generate exceptions to moral rules. But I think we can provisionally accept the conclusion that FH cannot be used to show that malicious acts are sometimes permissible.

Assessing Intrinsic Kantianism

Let us proceed on the assumption that FH leads to an intrinsic Kantianism that claims that malice, and perhaps some other motives, are strongly wrong-making. I think we then have enough material in front of us to evaluate the form of moral theory we thereby get. It is not acceptable. It is intuitively implausible, and, it indirectly undermines the largely objective nature of deontic status. I will close our consideration of Kantianism by contrasting FUL and FH.

Strongly Wrong-making Motives. The intrinsic Kantianism that we have found is generated by FH does not merely assert that malice can sometimes make an otherwise permissible action wrong. It asserts that acting maliciously is always wrong. I interpreted Kant's claim here to conjoin two assertions, *viz.*, that some actions are objectively wrong, and that any other actions malice leads to are also wrong. This conjunction 'proves too much'.⁵⁹ There are two ways to see the problem with the contention that there are strongly wrong-making motives.

First, we can think about it in terms of 'reflective equilibrium'.⁶⁰ I proposed to consider whether any moral theory could show that MM is true. Intrinsic Kantianism can do this. But surely this is only a necessary condition of theoretical adequacy. A theory should support MM in those cases where our considered moral judgments agree with it. There are cases where we believe that motives are wrong-making. But we do not believe that any particular motive is strongly wrong-making. We do not believe, therefore, that any malicious action is wrong. In the next chapter I will look closely at an example that

Kant himself discusses: punishing someone from revenge or malice. It may be that many actual cases of such an action are wrong because of their cruelty. (The execution of Saddam Hussein comes to mind.) But I will argue that if we present the right sort of imaginary case we believe that the punishment from malice there described is not wrong. Intrinsic Kantianism is therefore not fully in accord with our considered moral judgments.

There is a second reason to reject the claim that there are that are strongly wrong-making motives. The existence of such motives creates a fatal structural tension. Any acceptable moral theory must make deontic status a largely objective matter, and we have been able to interpret FH so that it seems to yield such a conception. But the idea that a motive like malice is always wrong-making cannot be reconciled with it. I noted in the last chapter that wrong-making motives could falsify PRC, which asserts that all obligations are objective. Now, it is possible that some actions are objectively obligatory and others are not. The problem with the notion of a strongly wrong-making motive is that it is inconsistent with the proposition that any action is objectively obligatory. Malice could favor telling the truth or keeping a promise or (I believe) not killing a person.⁶¹ This follows from Bentham's point about a single motive leading to various actions. So the idea of a strongly wrong-making motive allows for one objective deontic category only by undermining the possibility of another. A strongly wrong-making motive leaves room for objectively wrong actions, but then undermines the notion of objectively obligatory ones. Similarly, the notion of a strongly obligation-making motive leaves room for objective obligations, but then undermines the notion of objectively wrong actions.

It might seem to be impossible for an action like speaking the truth (or not killing) to be motivated by malice. After all, if FH establishes that an action is obligatory that seems to mean that it objectively furthers the good of the person who is owed the obligation. And a malicious agent makes a harm or loss to that person to be her end. However, there are in fact two ways in which malice could favor an objectively obligatory action. First, an act may in fact further the good of the person owed the obligation, but the agent may believe, falsely, that it harms her. An agent will always have some factual beliefs about how to realize the end that her motive establishes. For example, S may seek to produce pain to T as an end, and believe that the truth she speaks to T will produce pain for her. But, in fact, T may be grateful for the news. Second, it is actually possible for an act that does do harm to a person to be obligatory. One obvious example is punishment. It has been said that punishment is the only legally authorized act in which an agent of the government is expected to harm a citizen intentionally.⁶² If a punishing authority is indeed obligated to harm another person then this makes room for someone to carry out this obligation from malice or revenge. This is probably why an example of this form has been mentioned by philosophers since the time of Peter Abelard (c. 1140). We will consider it in the next chapter. There are other cases where malice could lead someone to perform an obligatory act. Whistle-blowing to the press and informing authorities about wrongdoing are examples. Violent acts in warfare are as well.

FUL and FH. The last two chapters have given us the opportunity to consider the controversial question of the relation between the first two formulations of Kant's Categorical Imperative, the Formula of Universal Law (FUL) and the Formula of Humanity (FH).⁶³ We have in effect done this by examining their treatment of some

actions stemming from bad motives like malice and racism. This is a different lens through which to look at the problem, I think. I did not employ exactly the same motive in test cases in the two chapters, although they are similar in a number of respects. (I used racism in the last chapter in part to bring out an important point about generating Reversed Role Processes (Reeps) in the universalization situation.) We saw that refusing to shake someone's hand from the requisite sort of racist ideology will pass the FUL test of permissibility. In this chapter we have seen that performing such an action from malice will fail the FH test. (A Kantian analysis of racism utilizing FH might show that performing it from racism would fail that test, too.) The two formulations are thus not equivalent, since they give different verdicts in fairly similar cases.

We can explain the difference in results by drawing on some remarks that I made before. FUL makes moral permissibility partly hinge on the effectiveness of the defensive Victim Protection Processes (Veeps) in the universalization situation. Since there are no fully effective Veeps with respect to hostile refusals to shake hands, or unprovoked insults, they pass the test. But they can certainly be seen as violating FH. It is a contingent matter whether rational agents (who are objective ends) can always thwart other agents who choose to 'act against' them. The other reason for the difference in results between FUL and FH stems from the role of Reeps. There are a number of actions which violate FH, but which do not disable a person from performing them herself, even if she is repeatedly subject to them. FH is, thus, 'stricter' than FUL. The verdicts that FH yields do not hinge on the issue of whether agents can prevent the affronts or aggressions that they are subjected to (in the universalization situation), or whether being subject to them disables an agent from performing them herself.⁶⁴

The Categorical Imperative is not a single principle that is formulated in different ways. It represents at least two principles. Neither of them, however, deals convincingly with the deontic status of actions from bad motives. The Formula of Universal Law and Formula of Humanity fail, in interestingly different ways, to give plausible support to MM. FUL judges acts from bad motives to be permissible in cases where we think they are wrong, and FH judges some of these motives to be strongly wrong-making. However, the most plausible position to take about the deontic relevance of a bad motive like malice is to say that it makes some, but not all, cases of acting from it wrong.

¹ G 96 (429). Emphasis deleted.

² However, the statement at G 104 (436), does mention maxims.

³ As I noted, Wood contends this in KET 97, and ch. 4, esp. pp. 139-41. For a response see Reath, op. cit., chs. 5 and 7, esp. pp. 160-1, n.7.

⁴ I have found the following interpretations of FH to be especially helpful: DPR, ch. 2; CKE, chs. 4 and 5; KET, ch. 4; Derek Parfit, Climbing the Mountain (manuscript available online), chs. 4 and 5; and Scanlon, Moral Dimensions, op. cit., ch. 3. Velleman, Self, op. cit., pp. 88-96 and 98-101 is suggestive.

⁵ VHK chs. 2-4.

⁶ G 96 (429) says the principle can be used to “derive all laws for the will”.

⁷ Wood accepts this, although his statement, “FH fundamentally demands...[that] our actions...express proper respect or reverence for the worth of humanity” might suggest otherwise. KET 141, his emphasis. See also 117; 363 nn. 7 and 8; 141-2; 147f.

⁸ Paton’s note at G 138-9; DPR 39-40; CKE 110; KET 118-9. VHK chs. 2-4 may not reject this general point.

⁹ G 57 (389); 92 (425). Wood notes that Kant is thus not a ‘speciesist’. KET 119-20.

¹⁰ G 101 (433). Italics deleted. Cp. G 96 (429).

¹¹ See G 105 (437); MetM 195 (392); DPR 38-41; CKE 110-114; KET 118-22. Dean, however, takes the essential feature of ‘humanity’ to be the Good Will, the concept introduced at G, 61-2 (393-4). He thus believes that the basic requirements of FH apply only to the treatment of rational agents who are committed to always acting permissibly (i.e., who have a Good Will). In theory this is a much smaller class of beings than the one that comprises beings able to rationally set ends for themselves. VHK chs. 2-4. In an earlier article Korsgaard seemed to endorse this reading. CKE 256-62.

¹² CKE 114. Korsgaard actually goes farther than this quotation suggests. She takes FH to be based on the view that “rational choice has...a value-conferring status.” That is, “rational choice makes its object good.” CKE 122. Dean takes the choices of a person with a Good Will to confer value. VHK122. Cp. Korsgaard’s earlier CKE 262.

-
- ¹³ Jonathan Bennett, The Act Itself (New York: Oxford University Press, 1995), p. 218. cp. Scanlon, Moral Dimensions, op. cit., ch. 3 sec.3.
- ¹⁴ Parfit, Climbing, op. cit., ch. 5, sec. 17. Cp. DPR 42.
- ¹⁵ Hill notes that G 97-8 (430) does not employ the notion of using a person simply as a means. DPR 42.
- ¹⁶ MetM 142 (333).
- ¹⁷ Parfit, Climbing, op. cit., ch. 5, sec. 16.
- ¹⁸ Ibid.
- ¹⁹ Wood reaches this conclusion by a direct examination of its wording. KET 143.
- ²⁰ First, in passages around the canonical formulation of FH (G 96 (429)) Kant makes statements that seem to be alternative versions of it that do not mention treating people as means. He writes, “Rational nature exists as an end in itself.” G 96 (429), emphasis deleted. Cp. G 98 (430-1). Second, two of the four examples that he immediately uses to illustrate the application of FH do not even mention treating people as means. Third, in the two examples where he does mention treating people as means Kant’s use of this idea is “a stretch”. VHK 130. For example, he says that “attempts on the ... property of others” use “the person of another merely as a means.” G 97 (430). But property crimes generally do not involve treating people simply as a means. If S steals T’s car she does not treat T herself simply as a means. If all of this is correct, we naturally wonder why Kant’s canonical formulation of FH incorporates the notion of using people as means. For some speculation see VHK 130; KET 143-4.
- ²¹ Scanlon, Moral Dimensions, op. cit., ch. 3, sec. 3-4, tries to isolate such cases.
- ²² E.g., G 96 (428).
- ²³ G 96 (428). Emphasis deleted. “Objective end” also occurs at G 98 (431). Cp. G 96 (429): “an objective principle”. Gregor, the translator of MetM, says that Kant’s usage there of ‘objective end’ is different. MetM 29 n. 4.
- ²⁴ G 105 (437).
- ²⁵ G 95-6 (428-9). For interpretation see CKE 114-24 (and cp. 256-62); KET 124-32; VHK, ch. 6. I noted that Korsgaard takes Kant to be saying not only that we judge what pursuits have value, we confer it on them; Dean says this about persons with a Good Will. See n. 12 above. For criticism of this claim, see Scanlon, Moral Dimensions, op. cit., ch 3, sec. 1; Donald Regan, “The Value of Rational Nature,” Ethics 112 (2002), pp. 267-91; Langton, “Objective,” op. cit., esp. pp. 174-85.
- ²⁶ G 95 (428). Emphasis deleted. G 97 (429); G 98 (430; 431).
- ²⁷ Sidgwick calls it “perplexing”. ME 390. See also the authors cited at KET 362 n. 4.
- ²⁸ I am here indebted to Andy Johnson. Wood denies that it is unusual in English or German. But he is prepared to ignore what dictionaries say. KET 116; 362 n. 4.
- ²⁹ J. L. Mackie, Ethics: Inventing Right and Wrong (Harmondsworth, England: Penguin, 1977), p. 24. Mackie says that Plato’s Forms are “a dramatic picture of what objective values would have to be.” Ibid., p. 40. But Kant’s notion of an ‘end in itself’ is just as dramatic. Of course, there are fundamental differences between Kant and Plato about the nature of the entity that has objective value. See KET 129-30. Cp. Langton, “Objective,” op. cit., esp. pp. 180-85.
- ³⁰ DPR 43-50, draws on a number of passages to provide “clues regarding what it means to treat humanity as an end.” (46) Note that Hill (47-50) takes an important clue to be

Kant's famous treatment of "dignity". Wood also does so. KET 115. But this passage actually occurs after other formulations of the Categorical Imperative (the Formulas of Autonomy and the Kingdom of Ends).

³¹ G 105 (437).

³² G 98 (431). Cp. G 96 (428); G 104 (436).

³³ This is how Herman interprets it. PMJ 13-15.

³⁴ I draw on DPR 50; PMJ 229.

³⁵ G 98 (430). Emphasis deleted. Similarly, when Kant discusses the duty to develop one's talents, he writes: "...it is not enough that an action should refrain from conflicting with humanity in our own person as an end in itself: it must also harmonize with this end." G 97 (430). Emphasis deleted.

³⁶ One would think that the application of FH to violence and killing is straightforward. But in fact there are a number of oddities in Kant's treatment of this topic. He is quite explicit in condemning suicide, and in appealing to FH to explain why it is wrong. He says that "I cannot dispose of man in my person by maiming, spoiling, or killing." G 96-7 (429). We expect him to say the same thing about killing or maiming other persons. But he does not. Herman says that Kant rarely mentions violence towards others. In addressing this gap in Kant's work she utilizes FUL, not FH, to establish a presumption against a maxim of violence to further one's self-interest. PMJ ch. 6.

³⁷ Hill finds a non-paternalistic benevolence in Kant's remarks. DPR 54-5. Cp. MetM 248 (454).

³⁸ Scanlon, alluding to the argument of G 95-6, argues that "it is much more plausible to say that the fact that something is some other person's end can give rise to a reason for us to promote it, or at least not interfere with it, than it is to say this in the case of our own ends." Moral Dimensions, op. cit., ch. 3., sec. 1. See note 25 above.

³⁹ In Kant's treatment of deception and "attempts on the freedom and property of others" he says that they violate FH on the grounds that the victim "cannot possibly agree with my way of behaving to him." G 97 (429-30). He does not explain what this phrase means. For some discussion see DPR 46; CKE 126-7; 137-40; Parfit, Climbing, op. cit., ch. 4. I noted above that Kant also implausibly asserts that "attempts on the ...property of others" use "the person of another merely as a means." G 97 (430). DPR 50-55 surveys the natural applications of FH, and Kant's own views. See also the survey in KET 140-1.

⁴⁰ Hill makes a similar point about "sharing the ends of others". DPR 45.

⁴¹ G 97 (429-30).

⁴² As I remarked above (note 11) Dean argues that 'humanity' in FH refers to persons who have a Good Will. This interpretation of FH differs significantly from the one I have discussed. Even if Dean's interpretation is correct, though, it does not alter the basic points we need for discussing MM. Dean's interpretation as stated means that we have duties of non-deception, non-violence, beneficence, etc., to a smaller group of beings than the other interpretation says that we do. But Dean still seems to conceive of these duties (as well as the failures to carry them out, etc.) in the objective way I explained in this section. Furthermore, Dean says that, for Kant, in practice the class of beings who are owed these duties roughly corresponds to the class of persons in the other interpretation. Among other reasons, this is because we can rarely be sure that someone does not have a Good Will. VHK 92-95. This means that the objective deontic categories of obligations,

permissions and prohibitions apply in practice to this larger class of beings. So we have an objective conception of deontic status in either case.

⁴³ MetM 251-9 (458-68). See Timmons, “Motive and Rightness,” *op. cit.*, pp. 279-85 on this material. For the role of motives see also Marcia Baron, “Love and Respect in the *Doctrine of Virtue*,” *Ibid.*, pp. 391-407, at 399-403; John Rawls, *Lectures on the History of Moral Philosophy* (Cambridge, Mass.: Harvard University Press, 2000), p. 187.

⁴⁴ Commentators have noted that when Kant comes to present a system of ethics in MetM, he largely employs FH to establish our duties. PMJ 133; KET, 139-41; CKE 124.

⁴⁵ In fact, Kant’s views on race are troubling, as Wood frankly admits. KET 3 and n.3.

⁴⁶ MetM 252-3 (460-1).

⁴⁷ Cp. Timmons, *op. cit.*, p. 287.

⁴⁸ In the last chapter I argued that the Practical Contradiction Interpretation of FUL makes motives relevant deontically only because of their effects. But we cannot assume that FH works the same way. The emphasis here on willing of course recalls the famous opening of G 61-2 (393-4). But now the character of a ‘willing’ is being used to establish a deontic status, not value or moral worth. Recall Herman’s assertion that “Kantian theory neither describes nor judges actions except as they are willed.” PMJ 91.

⁴⁹ MetM 253 (460).

⁵⁰ MetM 253 (461).

⁵¹ MetM 255 (463). Emphasis deleted.

⁵² MetM 251-2 (458). Emphasis deleted.

⁵³ MetM 251 (458).

⁵⁴ MetM 253 (461).

⁵⁵ MetM 253 (460).

⁵⁶ *Ibid.* Emphasis deleted.

⁵⁷ G 67 (399).

⁵⁸ It is interesting that Kant does not acknowledge the possibility of malicious acts of violence. See MetM 251 (458), quoted in the text above. I suspect that he thinks of malice this way because violent acts were implicitly treated earlier in MetM. But, as n. 36 above suggests, Kant does not explicitly discuss them even there.

⁵⁹ The problem here seems to be close to one noted by Jonathan Dancy. He says that Kantianism has no way to explain what he calls ‘contributory’ reasons; it has only an ‘overall’ notion of obligation. Jonathan Dancy, *Ethics Without Principles* (Oxford: Oxford University Press, 2004), pp. 67-70. In our discussion this manifests itself as follows: if acting from a certain motive as such is shown to conflict with FH, then such acting from that motive is always wrong (unless an appeal to FH itself can somehow be used to carve out an exception to this rule).

⁶⁰ TJ 48-51.

⁶¹ Timmons, “Motive and Rightness,” *op. cit.*, p. 283 claims that one cannot fulfill the Kantian duty of benevolence from malice. This is in tension, however, with the view that Kant accepts PRC. It is not clear whether Timmons thinks that, say, speaking the truth from malice is wrong in Kant’s theory.

⁶² Hart’s well-known definition includes the claim that punishment “must be intentionally administered.” Hart, *Punishment*, *op. cit.*, p. 5.

⁶³ For references to the scholarly debate on this point see Chapter 4, note 6, above.

⁶⁴ Korsgaard, in a discussion of Kant's argument for the impermissibility of lying to an intending murderer, says that FH is "stricter" than FUL, in the sense that FH finds all lying to be impermissible and FUL does not. CKE, ch. 5, at 151. I agree that FH finds more actions to be impermissible than FUL does. But I do not agree with Korsgaard's interpretation of these contrasting results, viz., that FH is a moral ideal suitable for a world where all agents carry out all their obligations, and FUL is the proper principle for governing a world where all agents do not carry out all their obligations. FUL finds even unprovoked acts of racist hostility to be permissible. These acts are not reasonable moral responses to the wrongful actions and intentions of others. Cp. CKE 100.