Kantian Moral Pessimism

"Those valiant men mistook their enemy, who is not to be sought in the natural inclinations, which merely lack discipline and openly display themselves unconcealed to everyone's consciousness, but is rather as it were an invisible enemy, one who hides behind reason and hence all the more dangerous. They sent forth *wisdom* against *folly* instead of summoning it against *malice* (of the human heart)."
>   -- Immanuel Kant, *Religion with the Boundaries of Mere Reason* (6:57)

"The Human Being is by Nature Evil"
>   -- *Religion with the Boundaries of Mere Reason* (6:32)

1. Introduction

Whether or not human beings are evil is not a popular topic among contemporary moral theorists.[1] Of course, one might rightly argue that ethics is fundamentally a normative discipline, asking about what *ought* to be the case, what the nature of obligation or virtue or goodness or justice is, not about whether human beings *in fact* live up to whatever normative standards there are. Moreover, one's normative theory does not, at least not directly, seem to imply any theory about whether or not human beings live up to the standards of that theory. If the primary focus of moral philosophy is on what proper moral norms are, it might naturally seem as though the empirical question of the extent to which people live up to these norms is at best a secondary concern.

Nonetheless, moral theories develop, whether intentionally or not, in the context of the problems and challenges that real human beings face in living their lives, and the moral commitments of human beings, for good or ill, structure these challenges. As a result, commitments regarding whether or not human beings are generally speaking good or evil can play widespread and often unnoticed roles in contemporary moral theorizing. In this paper, I raise just a few of the ways in which moral optimism – the view that human beings are generally good – affects contemporary ethical theory.[2] I start by looking at recent work by Gilbert Harman and (especially) John Doris, in which empirical studies of human moral experience play an important role in ethical reflection. Here, and wherever empirical work on actual human beings is taken to have normative implications, the issue of whether human beings are fundamentally good or evil is crucial for thinking about how to relate empirical studies to normative conclusions. If (and only if) human beings are basically good, then human behavior can provide a helpful and fairly straightforward guide for the way in which we ought to structure our moral theory. I then turn to a quite different context where moral optimism plays a role in ethical reflection, through examining its background influence in the work of Barbara Herman. Here I use Herman's account to make a broader point about how optimism informs the way that central issues in contemporary moral philosophy are discussed. Finally, I turn to Kant's own moral theory, focusing on his defense of a position that I label "moral pessimism."

Throughout this paper, the central contrast is between moral optimism and moral pessimism. Moral optimism is the view that human beings are basically morally good,

whatever this is taken to mean.[3] Moral optimists need not believe that human beings are omniscient nor that humans always do the best thing, but only that the main failings of most human beings are not primarily moral, but have to do with knowledge or competence or social conditions or (non-culpable) negligence or lack of self-control. Moral pessimism, by contrast, is the view that human beings are not basically morally good, that (at least) most human beings (at least) most of the time are, to some extent and in some way, *morally* deficient, and that at least a substantial portion of human misdeed is due to this moral deficiency.[4]

Moral optimism and moral pessimism are not, of course, the only possible options in assessing the status of human beings. Both views, as I have articulated them, depend upon a robust conception of morality, one that takes the notion of moral obligation and personal responsibility seriously. Nietzsche, for instance, is neither a moral optimist nor a moral pessimist; his optimism or pessimism lies "beyond good and evil."[5] Both also depend upon the notion that categories such as "morally good" can be applied to persons, rather than merely to actions or states of affairs.[6] Finally, and perhaps most importantly, the way that I have laid out the alternatives seems to leave out some important possibilities. In particular, both moral agnosticism – the view that one simply cannot know whether people are morally good[7] – and moral ambiguity – the view that most people are morally good in some respects and morally evil in other – seem to be more attractive (and more popular) alternatives to the black and white dichotomy between optimism and pessimism. There is certainly much to be said for these alternatives. Kant himself points out that "experience seems to confirm this middle position" (6:22).

Nonetheless, for the purposes of this paper I focus on optimism and pessimism, for three main reasons. First, my claim that commitments to claims about the moral status of human beings play a role in moral theorizing can be made drawing from any such claims, and the point is arguably more forceful when I can show a surreptitious commitment to a more extreme position (moral optimism) than to a more moderate one such as moral ambiguity or a more explicitly articulated one like a Nietzschean suspicion of morality itself. Second, the moral theory that leads Kant to his moral pessimism explicitly entails that moral ambiguity is not a genuine option: that is, the proponent of moral ambiguity either properly finds herself in the position of moral pessimism or illegitimately claims an ultimately unjustifiable, even if partial, moral goodness for human beings. (This claim will be backed up in more detail in section 4 below.) And, finally, it seems reasonable to read both agnosticism and ambiguity as sliding more towards optimism or pessimism in particular cases, and thus much of my discussion of optimism and pessimism will be relevant to the assessment of various sorts of agnosticism and ambiguity.[8] Thus in thinking about the question of whether human beings are, in general, morally good, I focus here on only two of many possible answers: Yes or No.[9]

2. *John Doris: Situationism, Virtue Ethics, and Optimism.*

Recently, both Gibert Harman and John Doris have invoked work in social psychology to argue against "character based virtue ethics" (Harman 2000: 176).[10] Their argument depends upon a wide range of empirical research designed to show that human behavior is determined primarily by one's situation rather than one's character, but two experiments are sufficient to highlight this general idea. One experiment, conducted at

Princeton Theological Seminary,[11] invited seminarians to participate in a study of religious vocation. Subjects began by filling out a questionnaire in one building and were then told to proceed to another building to give a short verbal presentation. After the questionnaire, subjects were told that they were either running late, right on time, or a little early for the presentation. Along the way, the subjects passed a person in extreme distress. Whether students stopped to help the person in distress correlated strongly with their level of hurry, with only 10% of the "high hurry" subjects stopping and 63% of the low hurry subjects stopping.

A second commonly cited experiment was conducted by Stanley Milgram at Yale.[12] In this experiment, subjects were instructed to administer electric shocks to a fellow participant (who was, in fact, an actor working with the experimenters) whenever that participant answered a question incorrectly on an exam administered by the experimenter. Although the subjects were told that the shocks were "painful but not dangerous," the shocks were labeled on a dial as increasingly severe, and the actor-participant expressed increasing levels of pain and anxiety, including statements such as "Get me out of here, please. My heart's starting to bother me. I refuse to go on. Let me out." Eventually, the actor screams hysterically and then becomes unresponsive; the subject is instructed to continue administering shock, taking unresponsiveness to be a wrong answer. At any point, if the subject expresses a desire to cease administering the shock, the administrator has a list of four things to say, as needed and in sequence: (1) "Please continue" or "Please go on." (2) "The experiment requires that you continue." (3) "It is absolutely essential that you continue." and (4) "You have no other choice, you must go on." If the subject persisted in asking to stop after being told these four things, he or she would then be excused. The results of this experiment (in several different trials and various iterations) is that all subjects were willing to administer at least "moderate shock" and two thirds of subjects continued to shock the fellow participant until the study was completed (at which point the dial reading was "XXX" and the subject was utterly unresponsive).

What are the implications of these studies for ethical theory? Both Harman and Doris use this empirical work to critique character based virtue ethics. The argument is fairly straightforward: What these experiments show is that human behavior is better explained by appeal to circumstances (whether or not one is in a hurry, whether or not an authority figure is present) than by character. Given these and other cases, in fact, "The experimental record suggests that situational factors are often better predictors of behavior than personal factors . . . . To put it crudely, people typically lack character" (Doris 2). Or, as Harman puts it, "There is no empirical support for the existence of character traits" (Harman: 178). The critique of virtue ethics then proceeds rather straightforwardly. Harman claims (and Doris defends in more detail[13]) "virtue ethics [of the sort both seek to criticize] presupposes that there are character traits of the relevant sort, that people differ in what character traits they have, and these traits help to explain differences in the way people behave" (Harman 168). If virtue ethics depends upon the fact that human character is primarily responsible for the way that human beings behave, and if situation *rather than character* is responsible for the way humans behave, then it looks like virtue ethics is empirically false.

Unfortunately, the above argument depends for its plausibility upon the

assumption of at least a moderate version of moral optimism. Doris and Harman need not be committed to the strong claim that most human beings are generally morally good, but they at least need to deny that most human beings are morally evil, and the strength of their argument really rests upon the general assumption of goodness in human nature.[14] A moral pessimist, looking at the data above, might read, not a refutation of the importance of character in ethical life, but a moral indictment of human beings. Rachana Kamtekar, in an important response to Doris and Harman's critique of virtue ethics, raises just this point:

> I do not doubt that the virtuous person would see her way to the right course of action, but perhaps there was no virtuous person among the subjects of these experiments: if virtue requires practical wisdom, *one would expect virtuous people to be rare*. (Kamtekar 485)

The Milgram experiments might show, not that human beings lack character, but that *these* human beings lack character, or that they have characters that are morally corrupt in ways that can be exploited by authority figures to get them to do extraordinary misdeeds for relatively trivial reasons.

This point can be made even stronger on a Kantian conception of the nature of moral character. For Kant, as for ancient virtue theorists, character is essential to moral life. Kant identifies "good character" with the "good will" (25:648) and claims, "The person that ought not trust himself with respect to his resolutions is in a state of hopelessness of all good" (25: 1387-88). For Kant, the character that is so important for having a good will is precisely the "stability and persistence in principles" (7:294) that social psychology calls into question. As Kant explains, "the most important part of character" is "that a human being has a constant will and acts according to it" (25:1386). However, even as Kant highlights the importance of character for the good will, he insists that it is *not* common among human beings: "the formal element of will as such, which is determined to act according to firm principles (not shifting hither and yon like a swarm of gnats), has something precious and admirable to it, *which is also something rare*" (7: 292, emphasis added, see too 6:651-52). Thus, he says, "before a good or evil character is built for a human being, a character altogether must be built, with which he first has a character in general, i.e., he first must get into the habit of acting from principles" (25: 630-31). And Kant does not think that this process of building a character is an easy task, nor one that can be accomplished quickly (see 7:294). For Kant, then, the Milgram and Princeton experiments provide quantitative confirmation of an empirical claim that Kant already affirms, an empirical claim that marks an important moral challenge for human beings. Kant does *not* take the fact that very few people act consistently in accordance with good principles to imply that consistent action in accordance with good principles is an inadequate moral ideal. Instead, he takes this fact to imply that moral virtue is an accomplishment, something that is, at best, rare.

Doris considers this sort of response to his interpretation of the phenomena, and he has responses ready at hand. Sometimes Doris (and Harman) explicitly invoke a moderate moral optimism to respond to interpretations of the phenomena that appeal to moral failings. Thus, with respect to the Milgram experiments, Harman asks (rhetorically), "can we really attribute a 2 to 1 majority response to a character defect? And what about the fact that *all* subjects were willing to go at least to the 300-volt level? Does *everyone* have this character defect?" (Harman: 171.) And Doris suggests, with

respect to a more troubling (because more real-world) case,

> virtually all Auschwitz doctors performed selections [deciding who would be killed and who would do forced labor]; did only men of bad character find their way to the camp?
>
> A persistent theme in accounts of the Holocaust is the perpetrators' 'ordinariness.' . . . . Unfortunately, it does not take a monster to do monstrous things. (Doris 54)

In the context, these can seem to be merely unjustified assertions of moral optimism, but Doris tries to make his particular optimism more palatable than the Nazi's—aren't—monsters quotation might suggest. As he explains, with respect to most of these experiments, "The problem that the empirical work presents is not widespread failure to meet heroic standards – perhaps this would come as no surprise – but widespread failure to meet quite modest standards" (Doris: 30).[15] Doris points out, in other words, that his inferences do not depend upon the claim that ordinary human beings are moral heroes, only that they are morally decent.

While making Doris's moral optimism seem more moderate and therefore more palatable, however, this response nonetheless simply highlights his commitment to moral optimism. And Kant's response, of course, is to argue that the data, rather than requiring a revision to moral theory, simply require abandoning even this moderate optimism. What the data show, in other words, is that ordinary human beings are not as morally decent as we think they are. But now, of course, the moral pessimist may seem merely to be stubborn. Surely the best explanation of why some seminarians stop and others do not cannot be the result of a thoroughgoing moral corruption on the part of those who do not stop, can it? Here is it important to note that a Kantian pessimist is not committed to the absurd proposal that Doris raises for explaining away these phenomena:

> It is not often going to be the case, as philosophers might be tempted to allege (see Feinberg 1992: 178), that those emerging as Failed Samaritans in some situation suffer a general "character flaw" while those presenting as Good Samaritans are motivated by a general surplus of benevolence. (Doris 39)

The point is not that, by sheer coincidence, characters always line up to make it seem like situations are doing the work. Instead, Kant *agrees* with Doris and Harman that the experiments show that the vast majority of people lack stable character traits. But for Kant, this lack is due to a widespread moral failing. It is not a lack that one must accept and build one's moral theory around. Lack of character is a *problem* that one must combat in order to bring about moral reform.

The difference between Kant's insistence upon the moral importance of character and Doris's dismissal of it has profound effects on the way each conceives of moral education. For Doris, "Rather than striving to develop characters that will determine our behavior in ways substantially independent of circumstance, we should invest more of our energies attending to the features of our environment that influence behavioral outcomes" (Doris 146). In fact, Doris even insists that there is substantial moral danger in excessive attention to character: "Many times a confidence in their character is precisely what puts people at risk in morally dangerous situations" (147). Kant's response, of course, is twofold. First, Kant would argue on purely normative grounds that Doris's program for moral education is precisely a program designed to lead people *deeper* into moral corruption. By deliberating avoiding situations that are morally difficult, human beings are capable of preserving their corrupt volitional structure while

becoming increasingly morally self-satisfied. Kant claims, in fact, that this is one of the primary strategies of impure and depraved wills. Doris rightly asks "which moral psychology [situationist or character-based] is better suited to effecting the practical aims of ethical reflection?" But Doris fails to appreciate the question of what, precisely, those ethical aims *are*. If Kant is correct that ethical reflection is precisely oriented toward the good *will* (rather than merely good actions[16]) – and this is something I have not defended here, but something that Doris says nothing to undermine – then Doris's program of moral education is a moral disaster.[17] Second, Kant often suggests, and some of the research Doris describes even confirms,[18] that attention to a fixed disposition of acting out of duty is precisely the best way to inspire human beings to emulate the life of virtue: "morality must have more power over the human heart the more purely it is presented" (5:156). Kant's attention to pure moral principles in his *Groundwork* and *Critique of Practical Reason* is oriented not merely towards philosophical clarity, but precisely towards illuminating the rigorous demands of morality in all its purity in order to inspire, through the sublimity of a life of moral virtue, "the greatest veneration and lively wish that [one] could become such a man" (5:156).

So whose interpretation of the data is more appropriate, Doris and Harman's use of it to argue against character or Kant's use of it to argue for a widespread moral failing in human beings? Social science cannot help us here, because the issue turns not on what the data show, but on how to read the normative *implications* of that data. That is, here the issue turns on one's moral theory. Kant's conception of moral virtue – even of the most ordinary kind – requires character, and he therefore reads the widespread lack of character as a reflection of a widespread lack of moral virtue. Doris and Harman are both unwilling to allow that moral corruption is widespread (at least, not extremely widespread) among human beings, so they read the widespread lack of character as a reflection of the irrelevance of character for moral virtue. But now the debate is just where Kant would want it to be (and in quite a different place from where Doris and Harman suppose it to be). For now, it should be much clearer that settling the dispute between Kant and Doris regarding moral optimism requires, *first*, doing some (pure) moral theory, that is, getting straight on whether it is possible to articulate any moral ideals that must be satisfied in order to call human beings good at all. Only once this task has been accomplished can one use empirical research to figure out how these ideals would apply to human beings and then the extent to which human beings actually live up to them.

One final point before leaving this section: it might just seem unfair to morally blame people for a lack of character, when even Kant exists that character is something that must be acquired over a long period of time. There is certainly an intuitive appeal to the claims of Doris and Harman; given how widespread the lack of character is, it seems better either not to hold people responsible at all or, as in the case of Doris, to develop an account of localized moral responsibility for particular actions in particular contexts given particular intentions (see Doris, chapter 7). One might respond to this concern in several ways. Kant, even though he claims that "the human being is evil *by nature*," nonetheless insists that one is evil "through one's own fault" (6:32). Fundamentally, Kant's way of reconciling these two claims appeals to the notion of transcendental freedom articulated in his Critical philosophy. The idea here is that however universal evil in human nature, each individual has this nature only because he or she freely

chooses it. One's free choice explains the nature that one has, not the other way around.[19] But one might also make some more particular points about the cultivation of character in particular. When Kant explains *why* character is so rare, he makes clear that this rarity is due to a sort of moral failing – reliance on instincts and inclinations –for which individuals can rightly be held accountable (7:294). Rachana Kamtekar has articulated a similar point (independently of and more lucidly than Kant). In the context of explaining situational variation with respect to deception, Kamtekar explains,

> It may require a strong interest (in the consequences of deceiving or not, or in the activity of deceiving or not) to lead one to extend one's strategies (or deception or non-deception) across situations . . . . [20] [T]he absence of a strong enough interest . . . may help to explain cross-situational inconsistency. (Kamtekar: 269-70)

The point here is that situationists tend to emphasize, in particular cases, that "the deeds in question do not require heroic commitment or sacrifice" (Doris 31). But what Kant and Kamtekar both point out is that the development of the sort of character that would act consistently across situations does require a sort of heroic sacrifice. The sacrifice is heroic not because it is impossible or requires an extreme form of moral fortitude but because the ordinary but categorical commitment to virtue required in order to cultivate and preserve character is, unfortunately, rare among human beings. Finally, even while saying that character is rare, Kant explains in his *Anthropology* a relatively simple set of guidelines that one can follow to cultivate and preserve character, including such things as "not to dissemble" and "to moderate our fear of offending against fashion" (7:294). In other words, Kant does not see character as innate, but he does see it as something that individuals can take responsibility for in themselves.

3. Barbara Herman: Optimistic Neokantianism

Both Doris and Harman write substantially outside of the Kantian tradition, so it is no surprise that they would differ in fundamental ethical commitments that could lead to profound differences in assessing the moral status of human beings. By contrast, Barbara Herman operates solidly within the tradition of Kantian moral philosophy. Herman fundamentally agrees with Kant's commitment to action from principle, and she even has highlighted the important place of character in a Kantian ethic.[21] Nonetheless, like Doris, Herman often operates under a background assumption of moral optimism. Unlike Doris, this background assumption does not play an essential role in Herman's key arguments. Instead, it helps set the contours of the debates in which she engages and the emphases in her responses to common criticisms of Kant. In this section, I take up three important issues with which Herman deals: the role of rules of moral salience in moral judgment, the importance of non-moral motivation in action, and the integration of morality into a coherent human identity. In each case, I show the role of moral optimism in shaping both the issues themselves and the way that Herman deals with them.[22]

Barbara Herman is perhaps best known for her work on rules of moral salience in the practice of moral judgment. For Herman, rules of moral salience "constitute the structure of moral sensitivity;" they "pick out certain aspects [of situations] . . . with the point of letting the agent see where moral judgment is necessary" (Herman 78). The idea here is that moral agents cannot simply apply a categorical imperative test to determine whether their maxims can be made universal. Agents must first formulate maxims, and, prior even to such formulation, must see their situations in ways that highlight morally

relevant features. As Herman explains,

> An agent who came to the CI [categorical imperative] procedure with no knowledge
> of the moral characteristics of actions would be very unlikely to describe his action
> in a morally appropriate way. Kant's moral agents are not morally naïve. In the
> examples Kant gives of the employment of the CI procedure (G422-423), the agents
> know the features of their proposed actions that raise moral concerns *before* they use
> the CI to determine their permissibility. (Herman 75)

A maxim that "when in financial distress, I will say to my friend that I will repay her, in
order to borrow money from her" seems quite universalizable; one who simply neglects
to notice the moral significance of the fact that one does not *intend* to repay will fail to
recognize the immorality of their actions. Similarly, one who fails to see his actions as
causing pain, or who fails to register the causing of pain as a morally significant feature
of a situation, will improperly apply the CI to her situation. One might see how this
might play out in something like a Milgram experiment. A subject who interpreted their
situation as one in which they were following the directions of the experimenter, but who
failed to attend to the apparent effects of those directions, would lack relevant rules of
moral salience.

So far, there is nothing in Herman's account that depends on or even suggests
moral optimism. The importance of rules of moral salience could even be interpreted in a
pessimistic direction, as just one more avenue for human corruption to manifest itself.
But Herman's discussion of these rules of moral salience includes three elements that,
together, suggest a substantial moral optimism in her account. First, Herman rightly
notes that on a Kantian account, human beings should not be directly held accountable
for acting in the light of bad rules of moral salience: "there seems to be no way to judge
actions apart from the way they are willed[, so] . . . morally defective RMS [rules of
moral salience] may not yield morally defective actions" (Herman 1993: 91). "It can be
permissible for agents with mistaken RMS to act in ways that would be judged
impermissible if their RMS were correct" (Herman 1993: 89). Of course, one *can* be
held morally responsible for bad RMS insofar as one is responsible for having the rules
of moral salience that one has (see Herman 89 fn 14). But, and this is a second important
element of Herman's view, although Herman allows that reflection can lead people to
change their rules of moral salience or to develop new ones, rules of moral salience are
"typically, . . . acquired in childhood as part of socialization" (Herman 78). Not only are
people not directly responsible for bad actions that are the result of bad rules of moral
salience, but people are typically not responsible for the rules of moral salience that they
have.

Even with these elements, Herman need not slip into moral optimism, but she
does when she introduces a third element of her picture: most misdeeds seem to be
ascribable to bad rules of moral salience, rather than to bad wills acting on good rules of
moral salience. Herman never makes this presupposition explicit, and she even points
out cases – such as Nazis (91-2) – where bad rules of moral salience are not morally
exculpatory. But her account is permeated with optimistic assumptions about the
prospects for improving behavior through improving rules of moral salience. Thus she
insists that "the scope of beneficent actions (how much good is done) *will be* greater for
persons who can more readily perceive the distress of others" (Herman 1993: 81,
emphasis added). And after saying that rules of moral salience "constitute the structure

of moral sensitivity," she explains that they not only "indicate when certain sorts of actions should not be taken without moral justification," but they can also "prevent certain kinds of action from occurring to the agent as real options for him (functioning here as a kind of moral taboo)" (Herman 78). And when describing the "Kantian agent," Herman not includes this agent's "responsiveness . . . shaped by moral knowledge (from the RMS)" but adds that "his attendant motivation includes a higher-order (or regulative) concern for the permissibility of his actions and projects" (Herman 83).

Combining the moral exculpability of bad rules of moral salience, the typical lack of responsibility for those bad rules, and the explanatory power of rules of moral salience in assessing behavior, one gets a picture of how moral optimism, though never explicitly articulated, permeates Herman's account. Most human misdeed are ascribable, not to evil wills, but to mistaken rules of moral salience. These mistaken rules are a moral problem and they should be changed (90), but the moral problem here is not a problem *with* the moral agent. Herman seems to have the view that most people are basically good, but that bad rules of moral salience lead people to do morally bad things.

The Kantian pessimist need not disagree with Herman's general account of rules of moral salience; instead a Kantian pessimist should take issue with one or more of the subsidiary hypotheses that lead to Herman's moral optimism. The first – that one is not directly responsible for one's rules of moral salience, such that one acting on bad rules of moral salience can nonetheless choose well – is linked to a fundamental aspect of Kant's moral theory, his commitment to the moral evaluation of *maxims* rather than (directly) of actions. This element of Herman's picture must remain in order for her account to be plausibly Kantian, and it helps to constrain and thereby clarify the nature of Kant's pessimism. Kant need not claim that every misdeed is the result of an evil will. Kant, like Herman, *can* allow for the possibility of misdeeds that are well motivated, for someone acting out of non-culpable ignorance of the morally salient features of their situation.

But Kant will disagree with both of Herman's other auxiliary hypotheses, if not strictly speaking, then at least by emphasis. It would, of course, be absurd to deny that one's upbringing can exert a substantial impact on the development of one's moral sensitivity (Herman's second emphasis). But where Herman *emphasizes* the role of social and cultural factors in the evolution of rules of moral salience, Kant would highlight the way in which deliberate *self*-corruption of one's rules of moral salience is a tool for moral self-satisfaction. For Kant, human beings cannot avoid moral self-judgment (6:438), and recognizing that one is morally evil brings a distinctive sort of dissatisfaction with oneself (6:394). At the same time, human beings are proficient at self-deception, especially in ways that can alleviate the pains of conscience. Thus one might seek to manipulate oneself so that one simply "fails to notice" areas in which the demands of the moral law would interfere with one's pursuit of one's non-moral ends.

Kant describes, for example, a case of "[self-]deceptive" inclination, where one's "natural inclination towards ease . . . makes [one] content with himself *when he is doing nothing at all* (vegetating aimlessly) because he *at least is not doing anything bad*" (7:152). The nature of this self-deception is such that one genuinely fails to notice as morally salient the fact that one is accomplishing nothing at all. The inclination to ease is not necessarily *stronger* than one's sense of duty (at least not in the moment). Instead, one employs a "ruse"[23] (25:503) whereby one does not even recognize one's duty

because one has warped one's rules of moral salience in the interests of one's inclination. For another example, one might imagine one in a privileged position in society who directs his attention away from structural injustices that would require a radical revision of his way of life. Even at the start, these redirections of attention need not be self-consciously immoral: "I have more important things to do," "This stuff is just too hard to figure out," "*I*'ve managed to work within the system, haven't I?" Over time, of course, such redirections of attention can warp one's rules of moral salience in ways that allow for the effective pursuit of personal happiness without moral qualm. Even more subtly, one might simply choose to spend more time with people whose rules of moral salience are more conducive to one's own non-moral ends. By choosing one's company carefully, one can subtly manipulate rules of moral salience to one's (nonmoral) benefit.[24] In all of these cases, what might seem to be innocent failures of RMS are actually due to self-deception in the service of one's bad will.

With respect to Herman's third point, the explanatory power of rules of moral salience, Kant again would point out that an emphasis on this explanatory power can often be a means for congratulatory self-deception. Kant points out that "we can never, even by the strictest examination, completely plumb the depths of the secret incentives of our actions," and he goes further, adding, "We like to flatter ourselves with the false claim to a more noble motive" (4:407). It is easy to see how this could play out in the context of rules of moral salience. Given Herman's claim, with which Kant would certainly agree, that "It can be permissible for agents with mistaken RMS to act in ways that would be judged impermissible if their RMS were correct" (Herman 1993: 89), one will have a natural tendency to reinterpret one's action in accordance with those rules of moral salience that make one's choices turn out to be permissible: "the human being knows how to distort even inner declarations before his own conscience" (8:270). This point is particularly worth emphasizing in the context of those who criticize Kant for misunderstanding the complexities of maxim formation. Kant was well aware not only of the possibility that the same actions could be represented under different maxims (a point that Herman rightly emphasizes), but also of the human tendency to use this fact as a means of justifying evil deeds.

The difference of emphasis with respect to the explanatory role of rules of moral salience in bad behavior also leads to significant differences between Kant and Herman regarding the proper emphasis of moral education. Kant is, of course, aware of the need to combat the tendency to develop bad rules of moral salience, and he suggests important measures for promoting better rules of moral salience, including "permissible moral illusions" such as politeness, which employ one's self-deceptive strategies against themselves,[25] as well as such things as seeking out "places where the poor who lack the most basic necessities are to be found" as a way of cultivating sympathy (6:457). But whereas Herman argues *against* a conception of Kantian education focused on "rational musculature" (Herman 1996: 43), Kant *defines* virtue as "strength of will" to act against one's inclinations in the light of the moral law as one sees it (6: 405). Thus for Kant, "a society . . . for the sake of laws of virtue" (6:94) will not only seek to improve moral sensitivity (in the sense of RMS), but will work to foster moral *strength*.

Moral optimism shows up in other of Herman's discussions as well. Arguably one of the most important issues in Kantian ethics is the role of non-moral motivations in

ethical life. In the *Groundwork*, Kant (in)famously writes that an act of beneficence
performed out of "an inner satisfaction in spreading joy" has "no moral content" and only
has "genuine moral worth" when performed "simply from duty" (4:398). For many
years, neokantians have been seeking to dull the force of this strong statement. Herman
takes up the issue in the context of a critique by Bernard Williams. As she summarizes
the critique,[26] it focuses on two central points:

> (1) Kantian morality often demands that we care about the wrong thing – about
> morality – and not about the object of our action and natural concern; (2) it leads to
> an estrangement from and devaluation of our emotions, especially in the rejection of
> emotions as morally valued motives. (Herman 24)

One who offers help to someone out of duty, rather than out of compassionate concern
for their well-being, seems both to care about the wrong thing and to mistakenly devalue
morally proper emotions. The case is particularly acute in the context of personal
relationships. In Williams's famous example, one who saves one's wife from drowning
because "it is morally permissible for him to save his wife" deliberates with "one thought
too many" (Herman 41).

Herman's response to Williams is, first, to distinguish between the "motive" for
an action and the "end" promoted by that action: "the end is that state of affairs the agent
intends . . . to bring about. The motive . . . is the way he takes the object of his action to
be good, and hence to be reason-giving" (25). In the case of helping another, one need
not have as one's *end* the fulfilling of the moral law; helping another person is not just a
means to fulfilling duty. Helping another is a means to the well-being of the other, and in
that sense, one who helps has a direct interest in the welfare of another. This direct
interest in the welfare of the other is present even in one who acts "out of duty," because
the role of duty here is *not* as a further *end*, but as the motive that makes the welfare of
another an end (in itself). Herman's second move goes even further. She argues that,
even when operating as a motive, one's duty very often functions only as a "limiting
condition" on other motives. That is, "it is not the function of the motive of duty to bring
about moral states of affairs." Rather, "it expresses the agent's commitment that he will
not act (on whatever motive, to whatever end), unless his action is morally permitted"
(31). So not only is duty not the end that one promotes in acting, it need not even be the
primary *motive*, at least not primary in the sense that it need not be the motive that
explains why an agent does this rather than that: "As a limiting condition, the motive of
duty in fact requires the effective presence of some other motive" (32). Thus when one
helps another, this help can legitimately have as its end the welfare of the other and can
even have as its motive one's compassion for the other. The motive of duty merely limits
this compassion to being effective only in cases where duty does not preclude its
expression.[27]

And Herman goes yet one step further, considering the case where "the motive of
duty . . . [is] sufficient by itself to bring the agent to do what is morally required" (32).
The case of beneficence may be a paradigm case of this, since human beings are
obligated to promote the welfare of others. And in these cases, Williams can raise a
further objection, that "the kind of help that can come from the motive of duty is not the
kind of help that is needed" so "it may be rational prefer an emotion-based to a morally
motivated action, [and so] it may be rational to place higher value on nonmoral than on
moral conduct" (Herman 33). Herman's response to this takes the notion of duty to the

level of character. What matters is whether or not one is a morally good *person*, not the moral worth of particular acts: "We probably will perform more acts with moral worth the better our will is. The number of morally worthy acts performed, however, is not proportional to the will's goodness" (35). For a person with a thoroughly good will, duty will be "ubiquitous" but not necessary "pervasive" (33) in the sense of being the primary motive for all of one's actions. In fact, Herman suggests, in cases where a beneficent action done from compassion will better promote the welfare of the person principally concerned, one with a good will and well cultivated compassionate feelings will act out of compassion. The fact that one has a good will means that duty is always present as a limiting condition, and that it functions as a primary motive when needed. But "it is not morally required that we always set the motive of duty between our feelings and our response to others[.] Kantian theory seems able to respect this aspect of our integrity as persons" (36-7)

As with her account of rules of moral salience, Herman's account here is one with which a Kantian pessimist need not strictly disagree. In this case, in fact, central aspects of Kant's account of human evil[28] fit quite well with Herman's account. Kant emphasizes that evil does not consist in "self-love" – the pursuit of one's contingent ends – but rather in "self-conceit" – the *unconditional* pursuit of these ends (5: 73). Kant goes further, arguing that one's non-moral ends have grounds in various *non-moral* predispositions *"to the good"* (6: 26-28).[29] And Kant even insists, in line with Herman, that a morally good agent is *not* one who eliminates these non-moral predispositions, but one who acts on them in such a way that they are subordinate to one's moral predisposition over the course of one's life (6:36). This conception of moral goodness encourages the idea that one can well act directly on one's non-moral desires, and such action expresses a morally good will as long as those desires are part of an overall structure that subordinates non-moral incentives to moral ones.[30] Finally, Herman's account of the proper role of inclinations fits well with Kant's claim in *Religion* that the inclinations themselves are not a threat to morality at all: "the ground of . . . evil cannot be placed, as is commonly done, in the . . . inclinations" (6:35).

Although Kant could well agree with Herman's account in these respects, however, Kant would add something important that Herman does not. Even if it is true that an idealized good human agent could act from non-moral incentives in a way that expresses a good will, when *real* human agents act from non-moral incentives, *we* do so from one or another form of *evil* will. As we will see in the next section, while Kant entertains the possibility of a person perfectly subordinating their non-moral incentives to their moral ones, he insists that *all* human beings *fail* to have the right volitional structure. And that means that Herman's hypothetical morally good will simply does not exist. While it is theoretically possible for a human being to act out of nonmoral emotions as motives while still preserving an overall guiding and limiting function to duty, and while this accurately characterizes what a good human will would look like, it does not characterize what actual emotion-guided human wills are like. Rather, human wills that are guided by non-moral incentives fall, for Kant, into one of three groups. Human wills are either frail (acting from non-moral motives despite a commitment to act according to moral ones), impure (doing what is morally correct but always only because cooperating sensuous incentives are present), or corrupt (explicitly and voluntarily

subordinating the moral law to sensuous incentives). Of these, the most important for responding to Williams's concerns are the frail and impure wills. For both of these sorts of wills, it is necessarily to cultivate precisely the virtue that both Williams and Herman discount, a capacity to act on the basis of principles rather than emotions, and in particular a need to strengthen the motive of duty vis a vis other motives, even those that are generally conducive to good actions. In the case of the frail will, the need here is fairly straightforward. The self-wrought human tendency to frailty requires the cultivation of a strength of will that can act on the basis of one's morally good maxims even in the face of strong inclinations to the contrary. Those who are frail *need* Williams's "one thought too many," because when the demands of justice or saving the lives of innocents require sacrificing the interests of a loved one, frailty can compromise good maxims.[31]

From the standpoint of Kant's analysis of the impure will, however, even more can (and should) be said about Williams's critique and Herman's response. Cases of impurity are precisely those where a commitment to the moral law and to inclinations that (in this case) concur with morality are present. Both Herman and Williams insist, in these cases, that there is nothing wrong with acting with inclination as one's immediate motive. Strictly speaking, Kant could agree, at least with Herman's formulation of this position. In a person with a properly ordered will, such that one acts on inclinations always only *because* those inclinations conform with the demands of morality, there is nothing wrong with allowing the immediate motive to action to be an inclination. And in cases – comforting a grieving loved one, expressing gratitude to a benefactor, making love to one's spouse – where acting immediately on the inclination is more conducive to morally required ends, one with a perfect will would act immediately on the inclination.[32] But in fact, human beings do not have perfect wills. Human beings have a tendency to subordinate the moral law to inclination, and one of the most common ways in which this is done is impurity – performing actions that are morally good, but doing so only when such actions also satisfy some inclination. Here one's underlying commitment subordinates morality to inclination, but the expression of that commitment may actually look *more* morally conscientious than the perfect will described above.

What will this mean, in practical terms? It will *not* mean a return to the picture of Kant according to which the only morally good actions are those done from duty alone. As I suggested, one who is impure may have an even stronger tendency to portray her actions as motivated from duty than one with a perfect will. But the recognition of impurity will have several important implications. First, while still cultivating those inclinations that are serviceable to morality, the emphasis of an agent concerned with impurity will be on the cultivation of moral "*strength*" (6:404) and moral *purity* – "the law being by itself alone the incentive, even without the admixture of aims derived from sensibility" (6:446). One's emphasis will be on ensuring that duty remains one's most fundamental and underlying commitment, even when one performs particular actions immediately from specific inclinations.

Second, one will be much more astute in one's moral self-assessment. Both Williams and Herman allow one to take self-satisfaction in performing good deeds from inclination, when one can see that such inclinations are part of a life of conformity with virtue.[33] But given the tendency to impurity in one's will, a Kantian agent will tend to discount a life of comfortable virtue. Such a Kantian agent, insofar as she is seeking to

grow in virtue, will certainly visit her sick friends and offer comfort to them. And insofar as such comfort is more soothing when offered "from the heart," she will allow her inclination to be the immediate motive to action. But she will not base her moral evaluations of herself on such actions, since it is impossible to distinguish reliably whether such an act with such motives constitutes progress towards virtue or regress deeper into impurity.

Finally, such a Kantian agent will seek out opportunities for virtues that truly test her moral resolve. This does not mean that she will seek to "despise [her friends] and do with repugnance what duties bids."[34] Rather, she will be particularly attentive to those occasions when her conscience pricks her with moral obligation while her inclinations rage against such duty. One who is morally self-satisfied with her generosity to her friends might be complacent in the light of these occasions, but a Kantian who is aware of the danger of moral impurity will recognize these cases as precisely those in which the rest of her life is brought into focus. Failing in duties that are difficult will not merely be an excusable weakness; it will taint all of one's easier exercises of virtue, suggesting that they are symptoms of impurity rather than moral excellence.

This account of different forms of evil in human nature, as well as Herman's account of the proper role of emotions in the life of a virtuous agent, both raise issues regarding the relationship between personal integrity and morality. And here, while Kant can again agree with the letter of Herman's response to Williams, his own response must be radically different. Williams raises the objection that Kantian morality "insists on dominion over even our most basic projects and intimate commitments, demanding a degree of attachment to morality that alienates us from ourselves and what we value" (Herman 24). We have already seen part of Herman's response above: Kant can respect our integrity at least so far as to allow that non-moral concerns can play a deep and even motivating role in the life of a Kantian good will. But Williams raises the ante on this objection by arguing that "in order to live at all, a person must have . . . 'categorical desires'" (Herman 37). Williams's objection is meant to be quite strong. As Herman explains,

> There is surely something true in the thought that our basic commitments and loves may such that they make us morally vulnerable: . . . we may find ourselves wanting to do something that impartial morality condemns . . . . But Williams wants to claim something stronger. Suppose our ground projects are what give us a reason to go on with our lives at all. Then if impartial morality can interfere with the pursuit of a person's ground project, there will be cases where an agent *could not* have reason to act as morality requires, for the *only* reasons we will have for acting are those that direct him to the impermissible pursuit of his ground project. . . . .So the Kantian idea that a rational agent will always have reason to act as morality requires is false. Since having ground projects is a condition of character . . . , the demands of impartial morality and those of character may conflict in deep ways. (37-8)

Herman's response to this objection is twofold. First, "[w]hile it is (psychologically) true that attachments to projects can be unconditional, it is not a requirement of the conditions of having a character that they be so" (39). Human beings need to have (nonmoral) commitments, but these commitments can be constrained by morality and still function as part of that sense of self that constitutes a character. Second, as there is nothing in proper

attachments that precludes them from being conditioned, "the moral agent is to be one who has a conception of himself as someone who will not pursue his projects in ways that are morally impermissible" (40). Morality itself, on this picture, *does* take the form of an unconditional attachment (one might even call it a "categorical desire"). And the morally good agent is one who has integrity, both in the sense of a set of projects, one of which is unconditional, and in the ordinary sense of integrity as moral restraint. In other words, "Kantian morality can be (and is meant to be taken as) defining a sense of self" (40).

Again, Herman is, strictly speaking, correct in her response to Williams. The life of a Kantian moral agent can be a life of integrity, and a commitment to morality can provide a sufficient governing commitment to provide a sense of self. (Kant provides a beautiful illustration of a life defined by its commitment to morality in the *Critique of Practical Reason*, 5: 155-6). But in this case, Williams is also certainly correct in his critique. For actual human agents, a commitment to morality *will* require deep conflict with the ground projects that give us a reason to go on with our lives. However theoretically possible it is to live a perfect life in which our identity is fundamentally (even if partially) defined by our commitment to moral virtue, such a life is not the life that any of us have actually chosen to live. The choice of our ground projects reflects a fundamental subordination of the moral law *to* these nonmoral projects, a subordination that is reflected in our individual choices and actions as well as in our deepest identity, the ultimate ground of these particular choices. As Kant explains,

> [Human] evil is *radical*, since it corrupts *the grounds of all maxims*; as a natural propensity, it is also not to be *extirpated* through human forces, for this could happen only through good maxims – something that cannot take place if *the subjective supreme ground of all maxims is presupposed to be corrupted*. (6:37, emphasis added, cf. 6:45).

For Kant, as for Williams, the result of the importance of nonmoral projects in human life is that an *unconditional* commitment to morality is at once impossible and irrational, and impossible precisely *because* it is irrational. Given that the choice to subordinate the moral law to particular inclinations and circumstances takes place at the level of the supreme ground of *all* of one's maxims, at what Williams and Herman might rightly call one's "identity," the decision to reverse this choice, to make the moral law unconditionally primary in one's life, just does not make sense. Given human corruption at the deepest level, Williams is absolutely correct that "our [nonmoral] ground projects are what give us a reason to go on with our lives at all[, so] . . . the demands of impartial morality and those of character may conflict in deep ways" (Herman 1993: 37-38).

Williams is even *almost* right in positing that "if impartial morality can interfere with the pursuit of a person's ground project, there will be cases where the agent *could not* have reason to act as morality requires" (Herman 1993: 37-8). Williams is correct that, from the standpoint of the agent's most fundamental commitments, the agent will not see the reasons given by morality as sufficient to trump those of her ground projects. But Williams is wrong in two respects. First, even if an evil agent does not see moral reasons as trumping, she will still see those reasons as reasons. Second, Kant maintains that even for corrupt moral agents, the moral law still has an authority that requires obedience. Even after agreeing with Williams that humans' categorical commitment to nonmoral projects (our "evil") is inextirpable (through human forces), Kant nonetheless

insists, "In spite of the fall, the command that we ought to become better human beings still resounds unabated in our souls; consequently, we must also be capable of it" (6:45). Like Williams and unlike Herman, Kant does not think that human beings can categorically choose morality without sacrificing their most fundamental ground projects, their identities, their characters. But like Herman and unlike Williams, Kant still maintains that human beings have a reason to categorically choose morality.

The implication of Kant's middle position here is that, for Kant (unlike both Williams or Herman), the life to which human beings are obligated is a sort of long, slow, painful suicide of one's deepest commitments. Kant explains the moral life of corrupt human beings as a "conversion," an "exit from evil and an entry into goodness, 'the putting off of the old man and the putting on of the new'" (6:74). As he insists, we must become "*other* human beings and not merely better human beings (as if we were already good but only negligent about the degree of our goodness)" (7:54). What is more, this transformation is deeply *painful*:

> The emergence from the corrupted disposition into the good is in itself already sacrifice (as 'the death of the old man,' . . .) and entrance into a long train of life's ills which the new human being undertakes . . . simply for the sake of the good" (6:74).

The "long train of life's ills" here can well be understood precisely in the terms with which Williams critiques Kant. What morality calls for really is a sacrifice of one's integrity, because this integrity has been constructed around a set of ground projects that have been given priority over the moral law.

This Kantian moral pessimism provides an account of the relationship between morality and integrity that is much truer to the real human experience of moral struggle than either Herman's or Williams's. With Herman, Kant insists that the demands of morality are humanly possible in the sense that it would be possible to live a coherent human life without non-moral *categorical* desires. Moreover, a morally virtuous life can accommodate non-moral desires, as long as these are kept in their place. In both of those senses, morality is consistent with human integrity, something that, in conscience if not in choices, most people affirm. But, with Williams, Kant acknowledges that living a moral life will involve pain beyond merely resisting temptations caused by important ground projects. The pain of living the moral life requires sacrifice of one's most deeply held commitments. In this sense, Kant helps show why Williams's critique is so appealing, so powerful, and – to a large extent – so correct. For those actually struggling to become morally better, Herman's easy ethical integrity is simply not true to real life ethical struggle. At the same time, Williams's complacent acceptance of categorical projects that can trump moral concern does not do justice to the morality that calls for such struggle. Kant's moral pessimism, in this case, seems to get it just right.

4. Kantian Moral Pessimism

Much of Kant's moral pessimism has already been elucidated by contrast with Doris and Herman, and further details are provided in the other excellent essays in this volume. In this final section, I offer some further elements in defense of Kant's pessimism. I start with Kant's argument for pessimism. I then briefly focus on how Kant's moral pessimism avoids four pitfalls that might seem implied by the view that "the human being is by nature evil" (6:32): the danger of undermining the principle that

ought implies can, the danger of moral despair, the danger of misanthropy, and the danger of complacency in the face of social and situational causes of evil.

First, Kant not only claims that human beings are evil, he also offers a detailed argument for this claim. Although he forswears a "formal proof" (6:32), Kant offers what is in effect a proof of moral pessimism. The bulk of this argument is found in Kant's *Religion within the Boundaries of Mere Reason*, but it begins with his approach to moral theory in the *Groundwork*. Two aspects of the argument of the *Groundwork* are crucial for Kant's eventual argument for moral pessimism. First, although Kant consistently affirms the principle that "ought implies can," he sharply distinguishes in the *Groundwork* between moral philosophy proper, which is a priori and purely normative, and moral anthropology, which considers the actual nature of human beings. Thus Kant neither derives nor even modifies his fundamental moral principle based on any facts about human nature, and he is therefore (unlike Doris) open to moral pessimism. The second key aspect of Kant's argument for pessimism that arises in the *Groundwork* is his insistence that good (human) wills must act in such a way that we do not "make an exception . . . for ourselves" (4:424). The moral law is universal and thus particular circumstances do not warrant an exception to it. Only one who acts out of a clear sense of the universal and exceptionless nature of morality truly acts "from duty."[35]

In *Religion*, Kant shows that his universalist ethic excludes what we previously called "moral ambiguity" – and what Kant calls "moral latitudinarianism" (6:22) – as an answer to the question of whether human beings are basically good or evil. In its place, Kant defends what he calls "moral rigorism," the view that there can be no intermediate between good and evil. As Kant explains,

> [I]f [a person] is good in one part [of his life], he has incorporated the moral law into his maxim. And were he, therefore, to be evil in some other part, since the moral law of compliance with duty in general is a single one and universal, the maxim relating to it would be universal yet particular at the same time: which is contradictory. (6:24-5)

Given that the moral law requires *unconditional* and *universal* compliance, those who act in conformity with what the moral law requires in some cases but not in others show that they never *really* make the true moral law the ground of their maxims, since any law the application of which depends upon particular circumstances cannot be the moral law.[36]

The rest of Kant's argument is fairly straightforward.[37] Certain actions cannot be willed in accordance with the moral law because they are directly contrary to right, and others are transparently based on maxims that are morally impermissible. In the performance of those actions, the moral law cannot be supreme. Given Kant's moral rigorism, the performance of such actions shows that the person who performs them does not make the moral law supreme in their lives, and is, thus, evil. Kant's argument then proceeds by highlighting "the multitude of woeful examples that the experience of human *deeds* parades before us" (6:32-3). In the present context, Kant could go even further, invoking the decades of quantitative research that Doris uses for different purposes; in Kant's hands, this research could clinch the empirical argument for radical evil in human beings.

Kant, of course, goes further than merely claiming that many human beings are evil; as noted, he finds evil in human nature itself. Kant does not offer many details

regarding the extension from widespread evil to evil as part of human nature itself, but his general account of anthropological knowledge allows for the legitimate inference from the empirical universality of a human trait to the (in principle revisable) ascription of that phenomenon to human nature generally.[38] What is potentially more problematic in the present context is the concern that such an ascription of evil to human *nature* undermines the notion that moral virtue is even *possible* for human beings. And if Kant's conception of moral virtue is literally *impossible*, this seems to be a serious blow to Kantian ethics. Arguably, in fact, much of the appeal of empirical arguments like Doris's is precisely that they seem to show that any moral philosophy that depends upon an ideal of stable character traits is simply beyond what humans can reasonably require of themselves. And Kant, of course, *is* committed to the principle that ought implies can: one rightly "judge[s] . . . that he can do something because he is aware that he ought to do it" (5:30).

So how does Kant reconcile the evil in human nature with the commitment to the possibility of acting morally? First and foremost, Kant resolves this apparent tension through his transcendental account of human freedom. In his *Critique of Pure Reason*, Kant argues that *all* empirical claims – including claims about human nature – are claims about appearances, and appearances are grounded in "things in themselves." And in his practical philosophy, Kant shows that human agents are free things in themselves that ground their appearances in the world. However one interprets these claims,[39] Kant's point is that human freedom is primary over even the most basic claims about human nature. In *Religion*, he specifically reiterates this point with respect to the radical evil in human nature, adding an important qualification about the sort of universality implied by ascribing evil to human nature:

> "He is evil *by nature*" simply means that being evil applies to him considered in his species; not that this quality may be inferred from the concept of his species (from the concept of a human being in general, for then the quality would be necessary), but rather that, according to the cognition we have of the human being through experience, he cannot be judged otherwise, in other words, we may presuppose evil as subjectively necessary in every human being, even the best. Now, since this propensity must itself be considered morally evil, hence . . . something that a human being can be held accountable for, . . . it must . . . always come about through one's own fault . . . ([be] brought upon us by ourselves). (6:32)

That human beings are evil by nature does *not* mean that it would be in principle *impossible* for a human being to be morally perfect, only that *no* human beings are *in fact* morally perfect. And Kant's first *Critique* shows precisely how this could be: empirically, one can reasonably deduce the universality of evil, but this universality is *ultimately* grounded, not in its empirical causes, but in the free choices of human agents.

At this point it is helpful to distinguish several different ways in which one might use the dictum that "ought implies can." Kant's primary use of this doctrine is to argue *from* obligation *to* possibility, as in the passage quoted above, where one "judges . . . that he can do something because he is aware that he ought to do it" (5:30). Another important use of the notion that "ought implies can" comes in the use of empirical data to specify the details of our obligations. Thus human beings are not obligated to feed every hungry person in the world because this is beyond the scope of our powers. Or, as for Kant, one might claim we are not obligated to *feel* love for others because "I cannot love

because I will to" (6:401, see too 4:399). And sometimes, these human limitations can even *expand* the scope of moral requirement: we are morally required to offer help to others in part because we need it ourselves, and we are morally obligated to be polite primarily because of moral failings to which politeness is a partial antidote. In these cases one proceeds from knowledge of the way in which human faculties relate to each other to construct a sort of physical-biological possibility that constrains (or sometimes expands) the particular duties that proceed from the moral law. Yet a third possible use of "ought implies can" goes further than this, and further than Kant. One might use empirical claims about human capacities to moderate the demands of morality *in general*. Here one argues that the moral law itself is not applicable to human beings, or not applicable in its purity, because of limitations in human nature. This is a use of "ought implies can" that Kant considers as an abstract but disastrous possibility (see *Critique of Pure Reason*, Bxxiii-xix) and that Doris employs implicitly throughout his work and occasionally explicitly, though in a moderated way (see especially pp. 112-4).

By articulating an account of human freedom that makes perfect virtue a possibility for human beings even if it is a possibility that is never actualized, Kant can avoid being forced into the implications of this third way of applying the principle that "ought implies can." Even Doris admits,

> Understood in the obvious way, as a requirement that ethical theories not demand the "strictly" (for instance, physically or logically) impossible, "ought implies can" does not further my cause: I've given no reason for thinking that the realization of virtue is strictly impossible. (Doris 112)

This concession is all that Kant needs, and Kant not only provides a detailed working out of the strict possibility of human realization of virtue, he is also willing to accept the moral pessimism that Doris is determined to reject. For Kant, human beings are all evil precisely because all human beings act in ways that conflict with a moral law that we *could* obey.

Still, even if Kant can respond to the strictly philosophical concern with the coherence of moral demands that are universally transgressed, there seems to be an important existential dimension to this concern that Kant's clever transcendental move does not alleviate. One who becomes convinced that he is morally corrupt at the deepest levels of his character, even (or perhaps especially) if he recognizes that this corruption is through his own fault, may collapse into a paralyzing sort of moral despair. This despair is especially likely given Kant's moral rigorism; there is nothing that a person can do to make it so that he will *always* obey the moral law, since he has *already* failed it. Kant raises precisely this concern in *Religion* (6: 72), and his response is twofold.

The first aspect of Kant's response to the existential problem of human evil is an appeal to a set of religious concepts: God, immortality, and (especially) grace. Kant emphasizes the idea that one's radical evil cannot be extirpated "through human forces" (6:37) and suggests, "Some supernatural cooperation is also needed to [a person] becoming good or better" (6:44, see too 27:294f).[40] This supernatural cooperation – God's grace in both enabling our transformation towards good and treating this transformation as sufficient to constitute true virtue – is accomplished over our immortal lives, through an everlasting struggle towards virtue. Although this appeal to supernatural forces may seem an excessive demand (and may even warrant, in the minds

of many, rejecting Kant's pessimism altogether), Kant makes his appeal more palatable in two important respects. First, Kant repeatedly emphasizes that this supernatural aid is inscrutable: of "supernatural assistance . . . we can have not the least cognition" (6:191, see too 7:43-4; 27:331; 28.2.2:1120-21, 1225, 1319). To avoid despair, one must believe that there is some way to be morally good, but practical reason will not condone any lenience in the interpretation of the moral law. Thus there must be some supplement for our failings, some supplement that reason does not fully specify. Second, and perhaps more importantly, Kant emphasizes that grace does not absolve one of the responsibility to actively promote one's own virtue: "Granted that some supernatural cooperation is . . . needed . . ., the human being must nonetheless make himself antecedently worthy of receiving it" (6:44).[41]

The second key aspect of Kant's response to the existential problem of evil is his affirmation of the enduring presence of what he calls the "predisposition to personality," "the susceptibility to respect for the moral law *as of itself a sufficient incentive to the power of choice*" (6:27). Human evil involves subordinating that moral predisposition to our non-moral ones, but human beings never eliminate the predisposition to personality itself:

> [T]here is still a germ of goodness left . . ., a germ that cannot be extirpated or corrupted . . . . The restoration of the original predisposition to the good in us is not therefore the acquisition of a *lost* incentive for the good, since we were never able to lose the incentive that consists in the respect for the moral law . . . . The restoration is therefore only the recovery of the *purity* of the law, as the supreme ground of all our maxims (6:45-6, see too 6:49; 7:43, 58-9)

Even when one subordinates the moral law to non-moral incentives, the force of the moral law is still felt, and anxiety over one's radical evil can itself be a sign that one's predisposition to the good is still present. And this means that however clear it is that one has made a commitment to subordinate morality, one still has the resources to recognize the supremacy of the moral law and even to act out of respect for it.[42]

Of course, one still might wonder "how it is possible that a[n] . . . evil human being should make himself into a good human being," but Kant points out that "since the fall from good into evil . . . is no more comprehensible than the ascent from evil back to good, then the possibility of this last cannot be disputed" (6:45). What is absolutely necessary in order to be morally good is that one have a basic capacity for action out of respect for the moral law, and all human beings – even the most evil – still have this capacity. The existential problem of moral despair comes when one confronts one's own free choice *not* to act out of this respect. But the solution to this problem lies in the use of one's free choice, and the way that human free choice is determined is incomprehensible. Thus as long as one still has the capacity to recognize the moral law as binding, one's free choice of evil does not provide sufficient warrant for despair about the prospects of obeying that law.[43]

Even if Kant can avoid the sort of moral despair that would come with an extreme and hopeless moral pessimism, though, Kant's pessimism might still seem conducive to a paralyzing gloominess in humans' interactions with each other. This gloominess could take at least two different forms. First, one might think that a Kantian moral pessimist will be inclined towards misanthropy, constantly fixed on the failures of human beings.

Second, one might think that Kantian pessimist will underestimate the importance of social and situational factors in generating bad deeds, and will thus be quietist about social and political changes that could make the world a better place. Fortunately, Kant's approach to human evil addresses both problems.

With respect to misanthropy, Kant recognizes the tendency for a realistic assessment of human evil to lead to misanthropy. He claims that, in many cases, "someone becomes a misanthrope due to the sensation of virtue, not because he despises people, but because he does not find them to be how he wants them to be" (25:553, see too 25:106, 530, 813, 932), and it might seem that Kant's own doctrine of radical evil would precisely be conducive to this sort of misanthropy. In fact, however, misanthropy for Kant is really caused by misplaced moral optimism. It is the disconnect between what one expects of others and what one finds that leads to misanthropy. If Kant is correct that people really are generally evil, then his moral pessimism inoculates against misanthropy rather than causing it.

Kant adds, at a strictly philosophical level, that what is worthy of respect (and even love) in human beings is *not* our perfect virtue, but our *capacity* for virtue. Although the good will is the only thing good "without qualification" (4: 393), Kant insists that "the human being an in general every rational being exists as an end in itself" (4:428); human beings, even the most evil, are worthy of respect.[44] Still, Kant recognizes that even if human beings strictly speaking deserve respect, one who is acutely aware of human evil may have difficulty bringing herself to this respect. For this problem, Kant recommends several specific measures. First, he argues that this propensity to misanthropy leads to a duty to avoid slander and even to construe the actions of others in the most favorable light.[45] One need not strictly deceive oneself about others' evil, but given one's tendency to allow attention to the evils of others to distract one from giving them the respect that they deserve, one ought to avoid excessive attention to others' moral failings: "People who always seek out the failings of [other] people become misanthropes." (This advice may be a particularly valuable warning for those who seek to apply Kant's account of evil to those around us.) Second, and relatedly, Kant offers strong endorsements of social interaction, and especially "polite" social interaction, in which, as he puts it, "Signs of well-wishing and respect, though originally empty, gradually lead to genuine dispositions of this sort" (7:152, see too 6:473-4). By acting politely, we both accustom ourselves to giving others the respect that they in fact deserve and we draw attention, both in ourselves and in others, to the more graceful and likeable qualities of human beings, rather than to the corruption that raises the specter of misanthropy. Finally, Kant suggests that the proper (pessimistic) way to think about human evil is precisely the opposite of what leads to misanthropy. Kant explains, "Misanthropy comes from a perverted concept of one's own importance and out of a black representation of [other] human beings" (25: 1364). In place of self-inflating attention to the evil in *others*, Kant generally directs his reflections on radical evil towards oneself (or towards a general humanity that includes oneself).[46] The result of Kant's pessimism is that at a rational, cognitive level, one recognizes both the radical evil of all human beings and the fact that all are nonetheless worthy of respect, while at the imaginative and affective level, one remains agnostic or even optimistic about others while deeply aware of the potential for self-deception regarding oneself.

Finally, though, even if Kant is able to reconcile his pessimism with misanthropy in general, there still seems to be a particular sort of social interaction that Kant's pessimism undermines and that the optimistic philosophies of both Doris and Herman seem well suited to promote. In particular, insofar as human misdeeds are largely the results of bad rules of moral salience or bad situations, these sources of ill encourage a sort of social struggle towards a better world. Human beings ought to work to cultivate virtue *in each other* through dialogue, moral inquiry, and social networks increasingly conducive to the good moral choices. By emphasizing human moral corruption, Kant might seem to undermine attention to these more external social causes of misfortune and arenas for moral improvement.

In fact, however, Kant's recognition of human evil leads him in precisely the opposite direction. In his *Religion*, Kant argues that even though human evil is self-wrought,

> the causes and the circumstances that draw one into this danger [evil] and keep him there . . . do not come his way from his own raw nature . . . but rather from the human beings to whom he stands in relation or association. (6:93)

The primary *means* by which we cultivate our own worst tendencies are social.[47] Competition, resentment, and a myriad of diverse desires arise only in social contexts, and the "passions," which for Kant are among the greatest hindrances to self-mastery and virtue, "assail his nature, which on its own is undemanding, *as soon as he is among human beings*" (6:94). Radical evil manifests itself not merely in individual wrongdoing, but in a social climate that fosters vice. The result, for Kant, is that one's struggle *against* one's self-wrought tendency to evil involves a struggle that is both oriented towards and dependent upon the reform of *society*: "The dominion of the good principle is not otherwise attainable . . . than through the setting up and the diffusion of a society in accordance with, and for the sake of, laws of virtue" (6:94). Rather than leading to a withdrawal into individual responsibility, Kant's account of radical evil leads him to a proactive social engagement oriented towards banding together to overcome our own evil tendencies. Like the suggestions of Doris and Herman, Kant's moral community will seek to improve people's circumstances in order to reduce temptations to vice, and it will seek improvement of rules of moral salience. But Kant's moral community will need to go beyond these measures, actively seeking to promote that character that can act consistently across situations and the moral resolution needed to act rightly in the light of one's rules of moral salience. In that sense, Kant sees the ethical life as even *more* socially engaged that do Doris and Herman; pessimism leads, not to disengagement, but to a more aggressive and better focused engagement in social, cultural, and political change.

5. Conclusion

Kant's view of the human species is not a particularly happy one. Kant not only believe that all human beings act wrongly, but that these bad actions are rooted in a fundamental failure of disposition. Human beings choose to subordinate the unconditional demands of the moral law to our shifting inclinations. As a result, we usually fail to act consistently, and insofar as we are consistent, such consistency is generally the result of our pursuit of one or another object of inclination. Although pessimistic, however, Kant's account of human nature is not *hopelessly* pessimist, nor

does it absolve human beings of responsibility for their evil, nor does it leave us inattentive to the role of circumstance, situation, or society in shaping our behavior. Instead, Kant's moral pessimism helps orient us to the real threats to morality in our world and thereby makes not only moral philosophy but also moral education and reform more relevant to the actual conditions in which human beings find themselves.

By accepting the presence of ubiquitous and pervasive evil in human life, Kant is able to put empirical research on human behavior to its proper use, not as a tool for revising the demands of morality, but as a way of showing in which areas – such as the cultivation of character – human evil must be combated. By diagnosing human evil in part in terms of self-deception, Kant can show how both situation-responsiveness and defective rules of moral salience are often not morally neutral facts about human beings but subtle strategies for preserving moral self-satisfaction while satisfying one's nonmoral interests. By recognizing the depth of human evil, Kant can explain the very real sense in which moral reform – even in the best cases – involves sacrificing what is deepest in one's self-conception; he thereby does justice to the real *struggle* that a life of moral hope involves. And by drawing attention to the way that society and situation facilitate the self-corruption of will, Kant can orient social reform and moral education in ways that get to the root of human corruption.

Throughout, Kant's attention to moral pessimism both allows for and requires a shift in emphasis in moral philosophy. Against Doris (and many others), Kant insists on keeping his accounts of the nature of morality free from attenuation in the light of facts about humans' actual behavior: "Any high praise for the ideal of humanity in its moral perfection can lose nothing in practical reality from examples to the contrary" (6:405-406). But against Herman (and many others), it is not enough for Kant to articulate this ideal of a human will in which moral virtue reigns supreme; Kant also insists upon the importance of a moral anthropology that highlights (and works to remedy) the pervasive evil that prevents human beings from realizing this ideal.

References

Adams, Robert. 1998. "Introduction" to *Religion within the Boundaries of Mere Reason*. Cambridge University Press.

Anderson-Gold, Sharon. 1986. "Kant's Ethical Commonwealth: The Highest Good as a Social Goal." *International Philosophical Quarterly* 26: 23-35.

------. 2001. *Unnecessary Evil*. SUNY Press

Darley, J. M. and C. D. Batson. 1973. "From Jerusalem to Jericho: A Study of Situational and Dispositional Variables in Helping Behavior." *Journal of Personality and Social Psychology* 27: 100-8.

Feinberg, Joel. 1992. *Freedom and Fulfillment*. Princeton, NJ: Princeton University Press.

Frierson, Patrick. 2007a. "Providence and Divine Mercy in Kant's Ethical Cosmopolitanism," *Faith and Philosophy* 24 (2007): 143-63.

------. 2007b. "Review of Richard Dean, *Kant and the Value of Humanity*," *Notre Dame Philosophical Reviews*.

------. 2006. "Kant's Empirical Account of Human Cognition." *Philosopher's Imprint*. 5.7: 1-36.

------. 2005. "The Moral Importance of Politeness in Kant's Anthropology," *Kantian Review* 9: 105-27.

------. 2003. *Freedom and Anthropology in Kant's Moral Philosophy*. Cambridge University Press.

Harman, Gilbert. 2000. *Explaining Value: and Other Essays in Moral Philosophy*. Oxford University Press.

Herman, Barbara. 1996. "Making Room for Character." *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*. Ed. Stephen Engstrom and Jennifer Whiting. Cambridge University Press.

------. 1993. *The Practice of Moral Judgment*, Harvard University Press.

Kamtekar, Rachana. 2003. "Situationism and Virtue Ethics on the Content of Our Character." *Ethics* 114: 458-91.

Korsgaard, Christine. 1996. *Creating the Kingdom of Ends*. Cambridge University Press.

Mariña, Jacqueline. 1997. "Kant on Grace: A Reply to his Critics." *Religious Studies* 33: 379-400.

Michaelson Jr, G. E. 1990. *Fallen Freedom: Kant on Radical Evil and Moral Regeneration*. Cambridge University Press.

Milgram, Stanley. 1963. "Behavioral Study of Obedience." *Journal of Abnormal and Social Psychology* 67: 371-8.

O'Neill, Onora. 1989. *Constructions of Reason: Explorations of Kant's Practical Philosophy*. Cambridge University Press.

Quinn, Philip L. 1986. "Christian Atonement and Kantian Justification." *Faith and Philosophy* 3: 440-462.

------. 1984. "Original Sin, Radical Evil, and Moral Identity." *Faith and Philosophy* 1: 188-202.

------. 1990. "Saving Faith from Kant's Remarkable Antinomy." *Faith and Philosophy* 7: 418-434.

Williams, Bernard. 1973. *Problems of the Self*. Cambridge: Cambridge University Press.

------. 1981. *Moral Luck*. Cambridge University Press.

Wood, Allen. 1999. *Kant's Ethical Thought*. Cambridge University Press.

------. 1991. "Unsociable Sociability: The Anthropological Basis of Kantian Ethics." *Philosophical Topics* 19: 325-351.

[1] Of course, there are some recent moral philosophers who take up this theme, and the present volume is an indication of its rising importance.

[2] This paper focuses only on two contemporary moral philosophers and only on a narrow range of issues. A full overview of the ways that moral optimism affects the topics and positions in contemporary moral theory would be worthwhile, but is beyond the scope of this paper. Among the further areas of influence, one would certainly have to mention theorizing about issues such as moral internalism, conflicts of duties, moral luck, and the relationship between morals and politics.

[3] As will be clear in my discussions of Doris and Herman, what it means to be morally good can differ widely between theorists, and some (such as Doris) are suspicious even of the notion that individual persons can be, on the whole, morally good or evil. For the purposes of this paper, it is not necessary to commit to a particular conception of moral goodness. One of the points of my paper, in fact, is that one's conception of moral goodness often can be surreptitiously informed by one's commitment to the empirical question of whether human beings are morally good. The way in which moral optimism plays out in quite different moral contexts will become much clearer in my application of the term to particular cases, below.

[4] As we will see, Kant thinks that *all* human beings are *radically* evil, but one need not take a position this strong to be what I call a moral pessimist.

[5] In a sense, of course, Nietzsche's pessimism about the herd-like nature of his contemporaries is a sort of revalued moral optimism; he thinks that most people are basically "morally good," but does not value moral goodness. In this paper, I am not interested in such alternatives to moral valuation.

[6] Doris explicitly calls this assumption into question (pp. 114-5), but his account nonetheless ends up at least being optimistic by exclusion; he denies moral pessimism, and in that sense takes human behavior as it is to set a standard for what human beings are capable of, and therefore what we are obligated to. One might, of course, expand the conception of moral optimism and moral pessimism to be less agent-centered, such that moral optimism is the view that generally speaking good actions are performed, or good states of affairs are brought about (or exist), while moral pessimism would say the opposite. In this paper, however, I focus on agent-centered optimism and pessimism.

[7] Kant's own pessimism, in fact, is deeply infused with moral agnosticism, so much so that many have taken Kant to *be* a moral agnostic. For a response to the claim of strong moral agnosticism in Kant, see Frierson 2003: chapter 5, where I argue that although Kant consistently denies the possibility of knowing that human beings are good, he allows and even insists upon the possibility of knowing that human beings are evil. Nonetheless, a further agnosticism arises from the way that Kant deals with the reality of radical evil in human nature. As we will see in section 4, Kant balances his moral pessimism with hope for moral improvement. And in this context, Kantian agnosticism reappears; one can never know with certainty whether or not one is making genuine moral progress. For Kant, this agnosticism has both philosophical roots (in his accounts of freedom, evil, and moral reform) and important practical benefits (as a means of warding off both moral despair and moral complacency). A full discussion of the implications of Kant's agnosticism is, however, beyond the scope of this paper.

[8] None of this is to deny the value of systematic study of the role of commitments to ambiguity and/or agnosticism in contemporary moral theory. This paper offers only the beginning of an analysis of possible influences that a commitment to a particular position on whether human beings are generally good can play in a moral theory. Further analyses of this sort would, I think, help bring these commitments more into the open.

[9] Moreover, for the purposes of this paper, I aim primarily to contrast Kant's moral pessimism with the moral optimism that is prevalent in much contemporary moral philosophy. For that reason, I focus in the next two sections on two particular examples of moral optimism in practice, before turning to Kant's moral pessimism. There are also, of course, important philosophers whose ethics is largely shaped by moral pessimism (Derrida and Levinas strike me as prominent examples of this, though a full defense of this claim would go beyond the scope of this paper), moral agnosticism, amoralism, and so on.

[10] There are, of course, important differences between Doris and Harman, but these do not impact the primary point of this paper.

[11] For details, see Doris 33-39. The original study is Darley and Batson 1973.

[12] For details, see Doris 39-51. The original research paper is Milgram 1963.

[13] See Doris 5-6, 15-22. For an important response defending the ancient virtue ethical tradition, see Kamtekar 2003.

[14] Both Doris and Harman take the claim that human beings are not generally evil to be a *conclusion* of their overall approaches to ethics. Harman suggests, for example, that his view will lead to "tolerance," saying,

> When things go wrong, we typically blame the agent, attributing the bad results to the agent's bad character. Even when things do not go bad, we are quick to interpret actions as expressive of character traits, often hostile traits. For example, a person with poor vision may fail to recognize an acquaintance, who then attributes this to coldness in that person. A greater understanding of the agent's situation and how it contributed to the action can lead to a greater tolerance and understanding of others. (Harman 177).

In fact, however, a "tolerant" view of others as basically good underlies the way in which both Doris and Harman use the empirical research to support their particular approaches to ethical theory.

[15] Elsewhere, Doris explains with respect to an experiment showing that people are likely to be more helpful after a small good fortune (finding a dime),

> Now one person did help, despite not finding a dime: perhaps the study shows only that compassionate people are few and far between . . . . But the cases I consider here, like the phone booth study, are ones where prosocial behavior looks to be 'minimally decent samaritanism' . . . ; the deeds in question do not require heroic commitment or sacrifice. (Doris 31)

[16] Doris repeatedly suggests that "ethical reflection is in the business of helping people behave better" (166), but he gives very little argument (for some, see pp. 15-20) to support this claim. By contrast, Kant gives extensive argument (see *Groundwork*, *Critique of Practical Reason*, and Book 1 of *Religion*) to support the claim that the structure of one's will, rather than one's actions per se, is the "business" of ethical reflection.

[17] Coming from a quite different direction, Rachana Kamtekar makes the same normative point:

> If virtue ethicists believe that saving a thousand Jews from concentration camps is sufficient to make Schindler virtuous, or must express a virtuous disposition, his war-profiteering and womanizing activities are a problem. But traditional virtue ethicists, at least, are not in the business of inferring virtues from actions, or even patterns of actions, no matter how admirable. From his actions alone it cannot be determined whether Schindler acted out of courage or, for example, guts and independence; if the latter, the traits are likely to be 'prosocial' or lead to 'prosocial' and admirable actions, only in certain circumstances, and there is no reason to expect that a man of guts and independence is either likely or unlikely ot be a war profiteer and a womanizer. (Kamtekar 486)

As in the case of Kant, ancient virtue theorists would not appreciate a program of moral education that left people's fundamental dispositions unchanged while constraining circumstances such that these dispositions led to better actions.

[18] See, for example, Doris, p. 50 ("obedience [in the Milgram experiment] was facilitated by perceptions of diminished responsibility") or p. 37 ("individual tendencies to accept rather than deny responsibility are positively related to a range of prosocial behavior").

[19] Famously, there are two quite different ways of explaining the relationship between human freedom and the world of appearances. One posits two distinct worlds that interact in a causal (or quasi-causal) way, such that one's non-empirical free choices cause the empirical world to be the way that it is. Another posits two distinct standpoints, such that one can think of the world empirically, in which case one ascribes one's actions to one's nature and circumstances, or practically, in which case one ascribes one's actions to one's freedom. For the purpose of responding to this concern about moral blame, either interpretation is adequate since neither account requires interpreting one's choice as dependent upon one's nature.

[20] In addition to the volitional component, Kamtekar adds, "and it may also be that transferring strategies from one situation to another that is quite different poses significant cognitive challenges." This cognitive component would, of course, alleviate responsibility for failing to develop character.

[21] See especially Herman 1996.

[22] These issues are discussed in Herman's *The Practice of Moral Judgment*, the text on which I focus in this section. There are, of course, other issues of importance where moral optimism informs Herman's account (such as conflicts of duties), but these are sufficient to give a general sense for the role of moral optimism in contemporary (neokantian) ethics.

[23] *eine List*

[24] Herman recognizes the possibility for culpable self-manipulation of rules of moral salience. She points out, for example, that "if an error is discovered but not incorporated in agents' maxims, there will be reason

to take their resistance as ground for moral criticism" (90). But Herman emphasizes the positive opportunities for improvement, while Kant would emphasize the myriad subtle ways in which human beings are capable of manipulating themselves to their moral detriment.

[25] See 7:151-153 and Frierson 2005.

[26] Throughout my discussion of Herman and Williams, I focus on Herman's discussion, and therefore I make extensive use of her summaries of Williams's critiques. For the original arguments in Williams, see especially "Morality and the Emotions" in Williams 1973 and "Persons, Character, and Morality" in Williams 1981.

[27] It is not the purpose of this paper to assess how well this reconstruction matches up with Kant's claims in the *Groundwork*.

[28] Strikingly, these elements do not appear in Herman's articulation of Kant's response. Granted, Herman is not seeking a primarily exegetical account of Kant here (see Herman, pp. 73, 93), but these core ideas still could have enriched her account, were she willing to endorse the pessimism with which they are associated in Kant.

[29] Strictly speaking, in fact, these are just three dimensions of a *single* predisposition to the good.

[30] See Herman 1996.

[31] Needless to say, frailty can also compromise moral demands for the sake of one's own interests, narrowly construed. But Kant (like Williams) sees cases where morality competes with our concern for loved ones as particularly important cases. In his account of how to inspire "the greatest veneration" for morality, he describes one who is morally resolved not to give false testimony against an innocent person but who is tempted to break this moral resolve: "represent his family, threatened with extreme distress and poverty, as *imploring him to yield* and himself, though upright, yet with a heart not hard nor insensitive either to compassion or to his own distress . . . yet remain firm in resolution to be truthful" (5:156).

[32] For a very helpful argument to this effect, see Herman 1993: 35-6.

[33] For Herman, this last qualification is essential; for Williams, it is not.

[34] From a well-known and oft-quoted satire by Schiller. See, for example, Wood 1999: 23.

[35] Note that this universal and exceptionless nature of morality need not contradict Herman's claim that a perfect moral agent might often act from non-moral incentives. The moral law must *always* function as a sort of limiting constraint on one's action, but non-moral incentives might acceptably be the immediate motives for action as long as they are situated within an overall volitional framework that subordinates non-moral to moral incentives.

[36] Kant's defense of this claim is based on the a priori arguments of the *Groundwork*. It is worth noting, though, that at least some of the empirical work to which philosophers like Doris appeal confirms Kant's sense that a moral middle ground is equivalent to moral corruption. As Doris notes, "attribution of negatively valenced traits may require very little in the way of behavioral consistency; perhaps one doesn't have to reliably falter, but only sporadically falter, to be counted a coward" (Doris 20, see too 97-98).

[37] There are, of course, important twists and turns in this relatively straightforward argument, and these have led to a great deal of controversy about precisely what sort of argument Kant aims to give here. I have discussed these in more detail in Frierson 2003. (The paragraph in which this note appears in a summary of my conclusions in Frierson 2003: 104-107.)

[38] Kant makes similar inferences throughout his work in anthropology and empirical psychology. For more discussion of this sort of "empirical universality" and how it contrasts with strict universality, see Frierson 2003: chapter 2. For contrasting treatments, see Allison 1990:154-5 and Wood 1999: 283-91.

[39] See footnote (13) xxx. For a defense of the possibility of explaining this grounding relationship in two-standpoint terms, see Frierson "Two Standpoints and the Problem of Moral Anthropology" (forthcoming xxx).

[40] See too 6:45, 61, 100-101, 143; 7: 43-44, 328; and 27:331. For more on grace in Kant, see Adams 1998, Frierson 2003:114-22, Frierson 2007a, Mariña 1997, Michaelson 1990, and Quinn 1986, 1984, and 1990.

[41] This first aspect of Kant's response to the existential problem of evil is made even more palatable when interpreted in the light of the second aspect. Kant connects the two in his *Conflict of the Faculties*:

> Grace is none other than the nature of the human being insofar as he is determined to actions by a principle which is intrinsic to his own being, but supersensible (the though of his duty). Since we want to explain this principle, although we know no further ground for it, we represent it as stimulus to good produced in us by God, the predisposition to which we did not establish in ourselves, and so, as grace. – That is to say, sin (evil in human nature) has made penal law necessary . . .; grace,

however, is the hope that good will develop in us – a hope awakened by belief in our original moral predisposition to good and by the example of humanity as pleasing to God in his son. And grace can and should become more powerful than sin in us (as free beings), if only we let it act in us or let our disposition to the kind of conduct shown in that holy example become active. (7:43)

One can thus interpret grace precisely to be the enduring possibility of acting in accordance with the moral law. In that context, although Kant makes use of theological concepts to articulate his doctrine of grace, what this grace amounts to can be incorporated into Kant's account of freedom. Just as human beings are undetermined by natural causes because of our potential to act in accordance with morality (our "predisposition to good"), we are undetermined even by our own (past) fundamental choices because this potential for moral motivation persists in us. Neither *potential* is due to our choice, so both are forms of "grace." But how one makes *use* of each potential is due to us, and so, on the one hand, we are evil, but, on the other hand, we have the responsibility to improve ourselves morally.

[42] Doris himself argues that "many Nazi war criminals are not straightforwardly understood as possessed of uniformly evil dispositional structures; much like Milgram's obedients, there is evidence that they experienced substantial conflict" (Doris: 54). The fact of psychological conflict is evidence that the Nazi's moral predisposition remained active, even as it was subordinated to situation-driven incentives.

[43] Finally, a third aspect of Kant's response is a consequence of the previous two. Although Kant is a moral pessimist in that human beings are without exception evil, he is also a philosopher of moral *hope*:

Assurance of [moral transformation] cannot of course be attained by the human being naturally, neither via immediate consciousness nor via the evidence of the life he has hitherto led . . . [y]et he must be able to *hope* that . . . he will attain to the road that leads in that direction" (6:51).

Kantian hope is essentially related to Kantian pessimism. There is no way to "naturally" gain assurance of moral goodness; in that sense, optimism is false. Still, "duty commands that [we] be good, and duty commands nothing that we cannot do" (6:47). This provides both a practical basis for *hope* – that is, one's obligations require that one have hope – and a rational (though insufficient) ground for hope, since the recognition of the authority of this command shows our enduring predisposition. The result is a fervent Kantian hope in moral goodness rooted *not* in any self-satisfied cognition of one's virtue, but in the recognition that even one's choice of evil has not excused one from obedience to the moral law. The result is a moral hope that is not *easy*: "He is a good human being only in incessant laboring and becoming, i.e. he can hope . . . to find himself upon the good . . . path of constant *progress* from bad to better" (6:48). With the cooperation of grace, one can hope for the best. But the best that one can expect is "an *endless progress* toward the complete conformity" of human wills with the moral law (5:122), a "battle . . . against the attacks of the evil principle" (6:93) and a constant "struggle" (6: 78) of one's efforts to be good against one's tendencies for evil. In place of moral despair, Kant offers a realistic but challenging moral hope.

[44] For further defense and explanation of this claim, see Frierson 2007b, Korsgaard 1996, and Wood 1999.

[45] "The intentional *spreading* of something that detracts from another's honor – even if . . . what is said is true – diminishes respect for humanity as such, so as finally to cast as shadow of worthlessness over our race itself, making misanthropy . . . or contempt the prevalent cast of mind, or to dull one's moral feeling by repeatedly exposing one to the sight of such things and accustoming one to it. It is, therefore, a duty . . . not to take a malicious pleasure in exposing the faults of others . . . but rather to throw the veil of philanthropy over their faults, not merely by softening our judgments but also by keeping these judgments to themselves; for examples of respect that we give others can arouse their striving to deserve it. – For this reason, a mania for spying on the morals of others . . . is by itself already an offensive inquisitiveness on the part of anthropology." (6:466)

[46] What is more, as we have already seen, this moral pessimism directed towards oneself is tempered with precisely the moral hope that can prevent potentially dangerous inwardly focused misanthropy.

[47] For more, see Anderson-Gold 2001, Frierson 2003, and Wood 1991 and 1999.