

Kantian Moral Pessimism

“Those valiant men mistook their enemy . . . They sent forth *wisdom* against *folly* instead of summoning it against *malice*.”

“The Human Being is by Nature Evil”

--Immanuel Kant, *Religion with the Boundaries of Mere Reason* (6:57, 32)

Whether people are evil is not a popular topic among contemporary moral theorists. Nonetheless, assumptions about whether humans are generally good or evil play widespread and unnoticed roles in moral theorizing. In this paper, I show some ways that moral optimism – the view that humans are generally good – affects contemporary ethical theory. I start with recent work by Gilbert Harman and John Doris, in which empirical psychology plays important roles in ethical reflection. Wherever empirical work is taken to have normative implications, the issue of whether people are fundamentally good contributes to thinking about how empirical studies relate to normative conclusions. I then turn to Barbara Herman’s work to show how optimism informs discussions of central issues in contemporary moral philosophy. I end with Kant’s “moral pessimism.”

Throughout, I primarily contrast moral optimism with moral pessimism. Moral optimists need believe neither that people are omniscient nor that they always do the best thing, but only that the main failings of most people are not primarily moral but have to do with ignorance or incompetence or social conditions or (non-culpable) negligence or lack of self-control. Moral pessimists, by contrast, believe that people are not basically good, that (at least) most people (at least) most of the time are *morally* deficient, and that many human misdeeds are due to moral deficiency.¹ Optimism and pessimism are not exclusive options for assessing people’s moral status. Both depend upon a robust conception of morality that takes moral obligation seriously. Nietzsche is neither a moral optimist nor a moral pessimist; his optimism or pessimism lies “beyond good and evil.” Both views also depend upon applying categories such as “morally good” to persons, rather than merely actions or states of affairs. Finally, I leave out alternatives like moral agnosticism (one cannot know whether people are morally good²) and moral ambiguity (people are good in some respects and evil in others). Here I focus on optimism and pessimism in part because my claim that commitments regarding people’s moral status play a role in moral theorizing is more forceful when I can show surreptitious commitments to a more extreme position (moral optimism) than to a more moderate one (like moral ambiguity) and in part because both agnosticism and ambiguity typically slide towards optimism or pessimism in particular cases, so discussion of optimism and pessimism is relevant to assessing other views.

2. *Situationism and Optimism.*

Recently, Gilbert Harman and John Doris have invoked social psychology against “character based virtue ethics.”³ They use empirical research that shows human behaviors determined by situation rather than character. For example, in the Milgram experiment,⁴

an experimenter got subjects to administer what they thought were deadly electric shocks to actors posing as fellow participants. Another experiment⁵ invited seminarians to participate in a study of religious vocation. Subjects filled out a questionnaire and were told to give a verbal presentation in another building. After the questionnaire, subjects were told that they were either late, on time, or early for the presentation. Along the way, the subjects passed an (apparently) extremely distressed person. Whether students stopped to help correlated strongly with their level of hurry, with only 10% of the “high hurry” subjects stopping and 63% of the low hurry subjects stopping.

Harman/Doris use such empirical studies to critique character based virtue ethics, claiming that they show that human behavior is better explained by appeal to circumstances (an authority figure present or being in a hurry) than by character: “The experimental record suggests that situational factors are often better predictors of behavior than personal factors . . . To put it crudely, people typically lack character.”⁶ Since “virtue ethics 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,”⁷ virtue ethics seems empirically false.

Unfortunately for Doris and Harman, this argument against virtue ethics depends for its plausibility upon moral optimism, at least to the extent of denying that most people are morally evil. A moral pessimist looking at the data might read, not a refutation of character’s importance, but a moral indictment of people: “[P]erhaps there was no virtuous person among the subjects of these experiments: if virtue requires practical wisdom, *one would expect virtuous people to be rare*.”⁸ This point can be strengthened given Kant’s conception of moral character. Kant identifies “good character” with the “good will” (VA, 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). The character so important for a good will is precisely the “stability and persistence in principles” (VA, 7:294) that social psychology calls into question. Kant explains, “the most important part of character” is “that a human being has a constant will and acts according to it” (VA, 25:1386). However, while Kant highlights character’s importance, he insists on its rarity: “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*” (VA 7:292, emphasis added, cf. 7:294, 25:630-1, MS 6:651-52). For Kant, the Milgram and Princeton experiments quantitatively confirm an empirical claim Kant already affirms. That few act consistently from good principles does not imply that consistent action is an inadequate moral ideal, but that moral virtue is an accomplishment that is, at best, rare.

In response to such interpretations, Harman sometimes explicitly invokes moderate optimism: “can we really attribute a 2 to 1 majority response to a character defect? . . . Does *everyone* have this character defect?”⁹ And Doris suggests, with respect to a more troubling 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? . . . Unfortunately, it does not take a monster to do monstrous things.¹⁰

Doris tries to make his optimism palatable by explaining, “[t]he problem 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.”¹¹ Doris’s argument does not depend upon the claim that ordinary people are moral heroes, only that they are morally decent. While making it more moderate and palatable, this nonetheless simply highlights Doris’s commitment to moral optimism. Kant can respond that the data, rather than requiring revision to morals, simply require abandoning even *moderate* optimism. Kant *agrees* with Doris/Harman that experiments show widespread lack of stable character traits. But for Kant, this lack results from widespread moral failing. Lack of character is not something to build a moral theory around, but a *problem* to combat in order to bring about moral reform.

The difference between Kant’s insistence upon character’s moral importance and Doris’s dismissal of it has profound effects on how 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.”¹² Against this, Kant first could argue on purely normative grounds that Doris’s program for moral education leads people *deeper* into corruption. By avoiding morally difficult situations, people preserve corrupt volitional structures while becoming increasingly morally self-satisfied. Doris rightly asks “which moral psychology is better suited to effecting the practical aims of ethical reflection?” But Doris fails to sufficiently defend what those ethical aims *are*. If Kant is correct that ethical reflection is oriented toward good *wills* (rather than good actions),¹³ Doris’s program of moral education is disastrous. Secondly, Kant argues, and some recent research confirms,¹⁴ that attention to a fixed dutiful disposition best inspires people to emulate the virtuous life (*KpV* 5:156). Kant’s focus on pure moral principles is not merely for philosophical clarity, but also to illuminate the rigorous, sublime moral law in order to inspire “the greatest veneration and lively wish that [one] could become such a [good] person” (*KpV* 5:156).

Social science cannot arbitrate between Kant’s and Doris’s interpretations of the data because these interpretations turn not on data but on its normative *implications*. Kant’s conception of ordinary moral virtue requires character, so widespread lack of character reflects widespread lack of virtue. Doris/Harman are unwilling to allow that moral corruption is widespread, so widespread lack of character must reflect character’s moral irrelevance. Deciding between Kant and Doris regarding optimism requires doing (pure) moral theory *first*, that is, getting straight on moral ideals. Only then can empirical research help one discern how ideals apply to people and the extent to which people actually live up to them.¹⁵

3. *Optimistic Neokantianism*

Unlike Doris/Harman, Barbara Herman operates solidly within Kantian moral philosophy. She agrees with Kant’s commitment to principled action and highlights character’s importance.¹⁶ Nonetheless, like Doris, Herman often operates under morally optimistic background assumptions. While optimism is inessential for Herman’s key arguments, this section examines three areas where optimism shapes her emphases: the role of rules of moral salience in judgment, non-moral motivation, and integrating morality with human identity.

Barbara Herman is best known for her work on rules of moral salience (RMS) in the practice of moral judgment. For Herman, RMS “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.”¹⁷ Moral agents cannot simply apply a categorical imperative (CI) procedure to determine whether maxims can be made universal. Agents must first formulate maxims by seeing their situations in ways that highlight morally relevant features.

An agent who came to the CI 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, the agents know the features of their proposed actions that raise moral concerns *before* they use the CI to determine their permissibility.¹⁸

The maxim “I will tell a friend that I will repay her in order to borrow money from her” is universalizable, but only because the borrower fails to include the morally relevant fact that she does not *intend* to repay.

No optimism so far. The importance of RMS could even be interpreted pessimistically, as another avenue for corruption. But Herman’s discussion of RMS includes three elements that together suggest substantial moral optimism. First, Herman rightly notes that Kantians should not hold people directly accountable for acting from bad RMS: “there seems to be no way to judge actions apart from the way they are willed[, so] . . . morally defective RMS may not yield morally defective actions.”¹⁹ “It can be permissible for agents with mistaken RMS to act in ways that would be judged impermissible if their RMS were correct.”²⁰ Of course, one *can* be held morally responsible for bad RMS insofar as one is responsible for having the RMS that one has.²¹ But, and this is a second element of Herman’s view, although reflection can provoke change, RMS are “typically, . . . acquired in childhood as part of socialization.”²² People are not directly responsible for acting from bad RMS, and people are typically not responsible for the RMS that they have.

Even these elements need not imply moral optimism without a third element of Herman’s account: most misdeeds seem ascribable to bad RMS rather than bad wills. Herman claims, for example, “the scope of beneficent actions . . . *will be* greater for persons who can more readily perceive the distress of others.”²³ The assumption seems to be that the primary reason for failures of beneficence (and other virtues) is a failure of RMS. Divergent emphases between Kant and Herman regarding the explanatory role of RMS are further reflected in different positions on moral education. Both Kant and Herman advocate promoting better RMS, for instance by visiting “places where the poor who lack the most basic necessities are to be found” (MM 6:457). But whereas Herman argues *against* education focused on “rational musculature,”²⁴ Kant *defines* virtue as “strength of will” (MM 6:405), and Kant’s “society . . . for the sake of laws of virtue” (R 6:94) not only seeks to improve moral sensitivity but works to foster moral *strength*.

Combining the moral exculpability of bad RMS, the typical lack of responsibility for those RMS, and their explanatory power in assessing behavior, one finds a pervasive but inexplicit moral optimism. Most human misdeed are ascribable not to evil wills, but to mistaken RMS. These mistaken rules are a moral problem and should be changed,²⁵ but they are not a problem *with* the moral agent. Herman seems to assume that people are basically good, but bad RMS lead to bad deeds.

Kantian pessimists need not disagree with Herman’s general account of RMS, but

only with one or more subsidiary hypothesis. The first – that one acting on bad RMS can nonetheless have a good will – is linked to Kant’s fundamental commitment to evaluation of *maxims* rather than (directly) of actions. This element of Herman’s picture is necessary for any plausibly Kantian account, and it helps constrain and thereby clarify Kant’s pessimism. Kant need not claim that *all* misdeeds are due to corrupt wills. Kant, like Herman, can allow that some misdeeds result from non-culpable ignorance of situations’ morally salient features.

But Kant disagrees with Herman’s other auxiliary hypotheses, if not strictly, at least by emphasis. It would be absurd to deny that upbringing exerts influence on moral sensitivity, but where Herman *emphasizes* this influence, Kant highlights how deliberate *self*-corruption of RMS is used to protect moral self-satisfaction. For Kant, people cannot avoid moral self-judgment (MM 6:438), and recognizing one’s moral evil causes self-dissatisfaction (MM 6:394). But people proficient at self-deception manipulate themselves to simply “fail to notice” areas where moral demands interfere with pursuing non-moral ends. Kant describes, for example, how one’s “[self-]deceptive” “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*” (LA 7:152).²⁶ Due to self-deception, one fails to notice as morally salient the fact that one accomplishes nothing. In the moment, the inclination to ease is not necessarily *stronger* than one’s sense of duty, but one employs a “ruse” (LA 25:503) that warps one’s RMS in the interests of inclination. Or one privileged in society might direct attention away from structural injustices that would require radical changes in one’s life: “I have more important things to do,” “This is just too hard to figure out,” “I’ve managed to work within the system, haven’t I?” Over time, such redirections warp RMS to allow for pursuit of personal happiness without moral qualm. For Kant, apparently innocent failures of RMS are often blameworthy forms of self-deception.

Regarding Herman’s third point, the explanatory power of RMS, Kant suggests that emphasizing morally-neutral causes of misdeeds is often a means of congratulatory self-deception: “[w]e like to flatter ourselves with the false claim to a more noble motive” (G 4:407). 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,”²⁷ people often reinterpret actions in accordance with RMS that make choices seem permissible: “the human being knows how to distort even inner declarations before his own conscience” (“Miscarriage,” 8:270).²⁸

Moral optimism thus plays important roles in Herman’s account of RMS, and given RMS’s centrality in her work, it is unsurprising that optimism shows up elsewhere. One important issue for neokantian ethics involves non-moral motivations in ethical life. In *Groundwork*, Kant (in)famously writes that acts of beneficence performed out of “an inner satisfaction in spreading joy” have “no moral content” and only get “genuine moral worth” when performed “simply from duty” (G 4:398). For many years, neokantians have sought to dull the force of this statement. Herman takes it up in the context of a critique by Bernard Williams. As she summarizes his twofold critique²⁹:

- (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.³⁰

One who helps from duty rather than compassionate concern for another's well-being apparently cares about the wrong thing and mistakenly devalues healthy emotions. The problem is particularly acute with 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" has "one thought too many."³¹

Herman's response to Williams involves first distinguishing between "motives" for actions and "ends" promoted by actions: "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 . . . reason-giving."³² When helping another, one's *end* need not be fulfilling the moral law but can be another's well-being. Direct interest in other's welfare is present even when acting "out of duty" because duty is not a further *end*, but the *motive* making another's welfare an end. Secondly, Herman argues that even as motive, duty often functions only as a "limiting condition." "It is not the function of the motive of duty to bring about moral states of affairs . . . [I]t expresses the agent's commitment that he will not act (on whatever motive, to whatever end), unless his action is morally permitted."³³ So not only is duty generally not one's end, it need not even be one's primary motive: "As a limiting condition, the motive of duty in fact requires the effective presence of some other motive."³⁴ When one helps another, one's help legitimately has the other's welfare as its end and can even have compassion as a motive. The motive of duty merely limits compassion to being effective only when its expression is not precluded by moral demands.

Herman goes further, considering cases where "duty . . . [is] sufficient by itself to bring the agent to do what is morally required."³⁵ Beneficence may be a paradigm case of this, since people are obligated to promote the welfare of others. In these cases, Williams objects, "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 to prefer an emotion-based to a morally motivated action [and so] be rational to place higher value on nonmoral than on moral conduct."³⁶ Herman's response takes duty to the level of character. What matters is whether one is a morally good *person*, not particular actions' moral worth: "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."³⁷ For a person with a thoroughly good will, duty will be "ubiquitous" but not necessarily "pervasive" in the sense of being the primary motive for all actions. In fact, Herman suggests, where beneficent action done from compassion better promotes another's welfare, one with a good will acts from compassion. A good will implies that duty is always a limiting condition and functions as 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."³⁸

As with RMS, Herman's account here is one with which Kantian pessimists need not disagree. In fact, aspects of Kant's account of human evil fit well with Herman's account. Kant emphasizes that evil consists not in "self-love" – the pursuit of contingent ends – but in "self-conceit" – the *unconditional* pursuit of these ends (*KpV* 5:73, cf. *R* 6:35). Even non-moral ends are grounded in predispositions "*to the good*" (*R* 6:26-8). And Kant insists, like Herman, that morally good agents do not *eliminate* non-moral predispositions but act on them in such a way that they are subordinate to one's moral

predisposition over one's whole life (*R* 6:36). This conception of moral goodness encourages the idea that acting directly on non-moral desires expresses a morally good will when those desires are part of a character that subordinates non-moral incentives to moral ones.³⁹

But while Kant could agree with Herman in these respects, he adds something important. An *idealized* good person could act from non-moral incentives and still express a good will, but when *real* human agents act from non-moral incentives, *we* express one or another form of *evil*. While Kant entertains the possibility of perfectly subordinating non-moral incentives to moral ones, he insists that *all* people *lack* this volitional structure; Herman's hypothetical good will does not exist. Human wills are frail (acting from non-moral motives despite moral commitment), impure (doing what is morally correct only because cooperating inclinations are present), or corrupt (explicitly subordinating the moral law to sensuous incentives) (*R* 6:28). Of these, the most relevant here are frail and impure wills. Combating frailty requires cultivating strength of will, something both Williams and Herman discount. Combating impurity raises even more complex issues. Impurity occurs when moral commitment coincides with inclinations. Herman and Williams insist that in such cases inclination can acceptably be one's immediate motive. Strictly speaking, Kant *could* agree. For properly ordered wills that act on inclinations *because* inclinations conform with morality, inclination could be an immediate motive. But people lack perfect wills. Impurity involves subordinating morality to inclination by performing morally good actions only because such actions also satisfy inclination. One's character subordinates morality to inclination, but the expression of that character may look morally conscientious.

Recognizing this danger need not imply that only actions from duty alone are morally good. In fact, *impure* wills may particularly portray their actions as motivated from duty alone. But recognizing the danger of impurity has important implications. First, combating impurity requires emphasizing moral "*strength*" (MM 6:404) and moral *purity* – "the law being by itself alone the incentive, even without the admixture of aims derived from sensibility" (MM 6:446). Second, one must be astute in self-assessment. Williams and Herman allow self-satisfaction in performing good deeds from inclination, but given the tendency to impurity, Kantian agents should typically discount comfortable virtue in self-assessment. Insofar as one seeks to grow in virtue, one will certainly visit and comfort sick friends. Since comfort "from the heart" is most soothing, one act from inclination as one's immediate motive. But one should not base moral self-evaluation on such actions since one cannot reliably distinguish whether they constitute virtue or impurity. Finally, one will seek opportunities to truly test moral resolve, not by "despising [friends] and doing with repugnance what duties bids,"⁴⁰ but by particular attention to occasions when conscience requires resisting inclinations. One morally self-satisfied with generosity to friends might be complacent on these occasions, but one cautious of impurity recognizes them as cases in which life is brought into focus. Failing in difficult duties is not merely excusable weakness; it taints easier good deeds, suggesting that they show impurity rather than virtue.

Herman's account of emotions and Kant's account of evil both raise issues about relationship between personal integrity and morality. Williams objects 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.”⁴¹ Here, while Kant can again agree with aspects of Herman’s response to Williams, his approach is radically different. Part of Herman’s response is integrity is preserved since non-moral concerns can play deep and even motivating roles in virtuous lives. But Williams deepens his objection by arguing that “in order to live at all, a person must have . . . ‘categorical desires’.”

There is surely something true in the thought that our basic commitments and loves may be 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.⁴²

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.”⁴³ People need to have (nonmoral) commitments, but these can function in constituting character even when constrained by morality. Second, because proper attachments can be conditioned by morality, “the moral agent is . . . one who has a conception of himself as someone who will not pursue his projects in ways that are morally impermissible.” Virtuous agents have integrity by having various projects subordinated under one unconditional one: morality itself. “Kantian morality can be (and is meant to be taken as) defining a sense of self.”⁴⁴

Strictly speaking, Herman is correct. Perfect moral agents’ lives would have integrity provided by governing commitments to morality. But Williams’s critique is also, for Kant, essentially correct. For actual human agents, morality *will* require conflict with ground projects that give us reasons to live. However theoretically possible a perfect life in which identity is defined by virtue, no human has actually chosen such a life. Our *actual* ground projects reflect fundamental subordination of morality to nonmoral goals, subordination expressed in individual choices and actions as well as in our deepest identity, the ultimate ground of these particular choices. Kant calls this deep-seated and categorical commitment to non-moral ground projects “radical evil.”

[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*. (R 6:37, emphasis added, cf. 6:45).

Because the choice to subordinate morality to inclinations occurs at the supreme ground of *all* one’s maxims, at what Williams and Herman rightly call one’s (deepest) “identity,” Kant agrees with Williams that “[nonmoral] ground projects . . . give us a reason to go on with our lives[, so] . . . the demands of impartial morality and those of character may conflict in deep ways.”⁴⁵

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."⁴⁶ From the standpoint of an agent's fundamental commitments, moral reasons seem insufficient to trump ground projects. But Williams is wrong in that even corrupt agents see moral reasons as reasons (albeit not overriding), and moreover, even for such agents, morality still *has* authority (even if unacknowledged) that *requires* obedience. While agreeing that humans' categorical commitment to non-moral projects ("evil") is inextirpable (through human forces), Kant nonetheless insists, "In spite of the fall, the command that we ought to become better people still resounds unabated in our souls; consequently, we must also be capable of it" (R 6:45). Like Williams and unlike Herman, Kant does not think that people 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 people have a reason to categorically choose morality.

The implication of Kant's middle position is that, for Kant (unlike Williams or Herman), moral life is a long, slow, painful suicide of one's deepest commitments. Kant describes such a life as "conversion," an "exit from evil and an entry into goodness, 'the putting off of the old man and the putting on of the new'" (R 6:74). We must become "*other* people and not merely better people (as if we were already good but only negligent about the degree of our goodness)" (SF 7:54), and this transformation is *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 (R 6:74).

The "long train of life's ills" is a sacrifice of a sort of integrity, one constructed around non-moral ground projects that were prioritized over morality.

Kant's pessimism about the relationship between morality and integrity captures valuable insights of both Herman and Williams. With Herman, Kant insists that morality's demands are possible in that a human life without non-moral *categorical* desires could have integrity. But with Williams, Kant acknowledges that moral life requires sacrificing one's deepest commitments. Kant shows why Williams's critique feels powerful and is even correct. Herman's easy ethical integrity is untrue to the real-life ethical struggle towards moral betterment, a struggle that really does involve giving up one's deepest commitments. But Williams's complacent acceptance of categorical projects that trump morals 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. Kant's Moral Pessimism

Having already elucidated much of Kant's moral pessimism by contrast with Doris and Herman, here I merely outline Kant's argument for pessimism and his response to four pitfalls that seem implied by the view that "the human being is by nature evil" (R 6:32). Kant argument for pessimism begins in *Groundwork*. First, *Groundwork* distinguishes between moral philosophy proper, which is a priori and purely normative, and moral anthropology, which considers humans' empirical nature. Because Kant neither derives nor modifies moral principles based on empirical facts, he is (unlike Doris) open to moral pessimism. Second, *Groundwork* emphasizes that good will do not

“make an exception . . . for [them]selves” (G 4:424). Because morality is universal, particular circumstances do not warrant exceptions to it. Only acting from exceptionless morality is acting “from duty.”

In *Religion*, Kant uses this universalist ethic to exclude what I previously called “moral ambiguity,” which Kant calls “moral latitudinarianism” (R 6:22). In its place, Kant defends extreme “moral rigorism,” denying any intermediate between good and evil.

[I]f [one] is good in one part [of life], he has incorporated the moral law into his maxim. And were he . . . 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. (R 6:24-5)

Because morality requires *unconditional* and *universal* compliance, acting in conformity with morality sometimes but not always shows that one never *really* makes the moral law one’s ultimate motive, since any law whose application depends upon circumstances cannot be the moral law.

Kant then argues for pessimism based on the fact that certain actions cannot be willed in accordance with duty because they are directly contrary to right, and others are transparently based on morally impermissible maxims. Given rigorism, those who perform such actions do not make duty supreme in their lives and are therefore evil. Kant then defends pessimism based on “the multitude of woeful examples that the experience of human *deeds* parades before us” (R 6:32-3). In the present context, Kant could invoke the quantitative research to which Doris appeals to clinch the empirical argument for human evil.

Kant’s argument then takes a turn that seems to raise serious problems. Kant not only argues that people are evil; he finds evil in human *nature*. While Kant’s account of anthropological knowledge allows for inference from the empirical universality of a trait to the (revisable) ascription of that trait to humans generally, the ascription of evil to human *nature* may seem to undermine the notion that virtue is even *possible*. If virtue is literally *impossible*, this seems a serious blow to Kant’s ethics. Arguably, much of the appeal of empirical arguments like Doris’s is due to the suggestion that any moral philosophy dependent upon stable character traits is beyond what humans can reasonably require of themselves. And Kant *is* committed to the principle that one is obligated only what to what one can in fact do (*KpV* 5:30).

So how does Kant reconcile evil in human nature with his commitment to the possibility of acting morally? First, 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 – refer to mere “appearances.” 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,⁴⁷ Kant’s point is that human freedom is primary over the most basic empirical claims about human nature. In *Religion*, he reiterates this point with respect to radical evil:

“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 . . . Now, since this propensity must itself be considered morally evil, . . . 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). (R 6:32)

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

In addition to this transcendental response to the problem of evil in human nature, it is important to distinguish several ways in which one might use the dictum "ought implies can." Kant primarily argues *from* obligation *to* possibility: one "judges . . . that he can do something because he is aware that he ought to do it" (KpV 5:30). One might also use empirical data to specify details of our obligations. Individuals are not obligated to feed every hungry person when this is beyond the scope of one's powers. And humans are not obligated to *feel* love because "I cannot love because I will to" (MM 6:401, cf. G 4:399). Sometimes, human nature *expands* moral requirements: we are obligated to help others in part because we need help and to be polite because politeness alleviates certain moral failings. In all these cases, empirical knowledge of physical-biological possibility helps specify particular duties.

Another use of "ought implies can" would go further, using empirical claims about human capacities to moderate the demands of morality *in general*, such as when Doris argues that morality is not applicable to people in its purity because of human limitations. Kant considers this use of "ought implies can" as an abstract but disastrous possibility (KrV Bxxiii-xix), and he avoids it by articulating an account of freedom that makes perfect virtue possible even if never actualized. Even Doris admits that he has "given no reason for thinking that the realization of virtue is strictly impossible."⁴⁸ This concession is all Kant needs. 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, people are evil precisely because we act in ways that conflict with a moral law that we *could* obey.

But even if Kant responds to *philosophical* concerns about universally transgressed obligations, there remains an important *existential* concern. One convinced of his own deep moral corruption, even (perhaps especially) if he recognizes that corruption is his own fault, may collapse into paralyzing moral despair. This despair is especially likely given Kant's moral rigorism: there is nothing that can make it so that one *always* obeys the moral law, since one has *already* failed.

In *Religion* (6:72), Kant raises and responds to this concern. The first aspect of Kant's response to the existential problem of human evil appeals to religious concepts: God, immortality, and (especially) grace. Kant emphasizes that evil cannot be extirpated "through human forces" (R 6:37) and suggests, "Some supernatural cooperation is also needed to [a person] becoming good or better" (R 6:44).⁴⁹ Although supernatural appeal seems excessive, Kant makes it more palatable by emphasizing the inscrutability of supernatural aid⁵⁰ and by insisting that grace does not absolve one of responsibility to

actively promote one's own virtue (R 6:44).

The second aspect of Kant's response to this problem 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*" (R 6:27). Evil involves subordinating that moral predisposition to non-moral ones, but people never eliminate it entirely:

[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 . . . is not therefore the acquisition of a *lost* incentive for the good . . . [but] only the recovery of the *purity* of the law, as the supreme ground of all our maxims (R 6:45-6; cf. 6:49; VA 7:43, 58-9)

Even when one subordinates moral to non-moral incentives, the force of morality is still felt; anxiety over radical evil even shows the enduring presence of one's predisposition to good. However clear one's subordination of morality, one still has resources to recognize its supremacy and act from respect for it.

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" (R 6:45). To be morally good, one requires a basic capacity for respect for the moral law, but all people 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 how such choice is determined is incomprehensible. As long as one has the capacity to recognize the moral law as binding, prior evil choices do not warrant despair about prospects for obeying that law.

A third aspect of Kant's response follows from the previous two. Although Kant is a moral pessimist, he is also a philosopher of moral *hope*:

Assurance of [moral transformation] cannot of course be attained by the human being naturally, . . . [y]et he must be able to *hope* that . . . he will attain to the road that leads in that direction" (R 6:51).

Hope is essentially related to pessimism. One cannot "naturally" be assured of moral goodness, but "duty commands that [we] be good, and duty commands nothing that we cannot do" (R 6:47). This fervent Kantian hope in moral goodness, rooted *not* in self-satisfied cognition of virtue but in recognition that even evil does not absolve one of responsibility, is not *easy*: one "is a good human being only in incessant laboring and becoming" (R 6:48). By grace, however understood, one can hope for the best. But the best is "an *endless progress* toward complete conformity" with the moral law (KpV 5:122), a "battle . . . against the attacks of the evil principle" (R 6:93) and a constant "struggle" (R 6:78) to be good against self-wrought evil tendencies. In place of moral despair, Kant offers a realistic but challenging moral hope.

Even if it avoids hopeless despair, however, Kant's pessimism might seem conducive to gloomy misanthropy focused on others' failures. Kant does, in fact, recognize that realistic assessments of evil can cause misanthropy: often "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" (VA 25:553, cf. 25:106,

813, 932). But misanthropy is caused by misplaced *optimism*, a disconnect between expectations and reality. If people are evil, pessimism inoculates against misanthropy rather than causing it. Moreover, what is worthy of respect is *not* perfect virtue, but the *capacity* for virtue. Although good will is the only thing good “without qualification” (G 4:393), “the human being . . . exists as an end in itself” (G 4:428); even the most wicked are worthy of respect.⁵¹ Given the difficulty of respecting those known as evil, Kant recommends avoiding slander and construing others’ actions favorably (MM 6:466). Without deceiving oneself, one can avoid excessive attention to others’ failings. Relatedly, Kant endorses “polite” interaction in which “signs of well-wishing and respect . . . lead to genuine dispositions of this sort” (VA 7:152, cf. MM 6:473-4). Through politeness, we accustom ourselves to giving others respect, and we draw attention to others’ likeable qualities. Finally, Kant suggests that the proper (pessimistic) way to think about evil is precisely the opposite of what leads to misanthropy: “Misanthropy comes from a perverted concept of one’s own importance and out of a black representation of [other] people” (VA 25:1364). In place of self-inflating attention to the *others’* evil, Kant directs such attentions towards oneself. At a rational, cognitive level, one recognizes both the radical evil of all 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 cautious about optimistic self-deception regarding oneself.⁵²

Finally, even if Kant separates pessimism from misanthropy in general, pessimism seems to undermine valuable sorts of social interaction; recognizing human misdeeds that result from bad RMS (Herman) or bad situations (Doris) encourages social struggle towards a better world. People can cultivate better virtue *in each other* through dialogue, inquiry, and social networks conducive to good choices. By emphasizing personal corruption, Kant seems to undermine such arenas for moral improvement.

In fact, however, Kant’s pessimism has the opposite implication. Although evil is self-wrought,

the causes and the circumstances that draw one into this danger and keep him there . . . do not come . . . from his own raw nature, so far as he exists in isolation, but rather from the people to whom he stands in relation or association. (R 6:93)

The primary means by which one cultivates one’s worst tendencies are social.⁵³ Evil manifests itself not merely in individual wrongdoing, but in the cultivation of vice-conducive social climates. Hence for Kant, struggle against self-wrought evil tendencies involves 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” (R 6:94). Rather than leading to withdrawal into individual responsibility, Kant’s pessimism leads to proactive social engagement. Kant’s moral community seeks to improve RMS (like Herman) and reduce circumstances that tempt to vice (like Doris), but it also goes further, actively promoting consistent character and moral resolution to act rightly in the light of one’s RMS. In that sense, Kantian ethical life is even *more* socially engaged than Doris and Herman’s proposals. Pessimism leads, not to disengagement, but to aggressive and focused engagement in social-cultural-political change.

5. Conclusion

Kant's view of the human species is not particularly happy. Bad actions are rooted in fundamental failures of character. People choose to subordinate unconditional demands of morality to shifting inclinations. We usually fail to act consistently, and what consistency we have usually results from pursuing nonmoral grounding projects. But Kant is neither *hopelessly* pessimistic nor absolving of people's responsibility to deal with evil. Instead, Kant's pessimism orients us to real moral threats and thereby makes both moral philosophy and moral reform more relevant to actual conditions of human life.

By accepting the ubiquity of evil, Kantian pessimists can use empirical psychology not to revise moral demands but to show where evil must be combated. By diagnosing evil partly as self-deception, Kant shows how often blaming evil on situation-responsiveness or defective RMS are subtle strategies for preserving moral self-satisfaction while satisfying nonmoral interests. By recognizing evil's depth, Kant does justice to the real *struggle* of a life of moral hope by showing why moral reform requires sacrificing one's deepest self-conception. And by drawing attention to how society and situation facilitate self-corruption, Kant orients social reform.

Kant's pessimism allows and requires a shift in emphasis in moral philosophy. Against Doris, Kant keeps accounts of morality's nature free from attenuation by facts about human behavior: "Any high praise for the ideal of humanity in its moral perfection can lose nothing in practical reality from examples to the contrary" (MM 6:405-406). But against Herman, Kant does not merely articulate an ideal of a morally good human will; he insists upon a moral anthropology that highlights and works to remedy the pervasive evil that prevents people from realizing this ideal.

¹ Kant thinks that *all* people are *radically* evil, but a moral pessimist need not take such a strong position.

² Kant's pessimism is so infused with agnosticism that many see Kant as morally agnostic. In response, see Patrick Frierson, *Freedom and Anthropology in Kant's Moral Philosophy* (Cambridge University Press, 2003).

³ Gilbert Harman, *Explaining Value: and Other Essays in Moral Philosophy* (Oxford University Press, 2000), 176.

⁴ See John Doris *Lack of Character: Personality and Moral Behavior* (Cambridge University Press, 2002), 39-51; Stanley Milgram, "Behavioral Study of Obedience," *Journal of Abnormal and Social Psychology* 67 (1963).

⁵ See Doris *Lack*; J.M. Darley and C. D. Batson, "From Jerusalem to Jericho: A Study of Situational and Dispositional Variables in Helping Behavior," *Journal of Personality and Social Psychology* 27 (1973): 100-8.

⁶ Doris *Lack* 2, cf. Harman *Explaining Virtue* 178.

⁷ Harman *Explaining Virtue*, 168. Cf. Doris 5-6, 15-22. For defense of ancient virtue ethics, see Rachana Kamtekar, "Situationism and Virtue Ethics on the Content of Our Character," *Ethics* 114 (2003): 458-91.

⁸ Kamtekar, "Situationism," 485.

⁹ *Explaining Virtue*, 171

¹⁰ *Lack*, 54.

¹¹ *Lack*, 30.

¹² *Lack*, 146.

¹³ Doris claims, "ethical reflection is in the business of helping people behave better" (*Lack*, 166) but gives little argument for this (cf. 15-20). For Kant's extensive argument that the structure of one's will, rather than one's actions, is the "business" of ethical reflection, see *G*, *KpV*, and *R*.

¹⁴ See Doris *Lack*, 50 ("obedience [in the Milgram experiment] was facilitated by perceptions of diminished responsibility") or 37 ("individual tendencies to accept rather than deny responsibility are positively related to a range of prosocial behavior").

¹⁵ One final point: it might seem unfair to blame people for lacking character when even Kant admits character must be acquired over a long period of time. Given the widespread lack of character, it might seem better either not to hold people responsible at all or to develop accounts of localized moral responsibility (Doris, *Lack*, chapter 7). However, even as Kant claims that "the human being is evil *by nature*," he argues that one is evil "through one's own fault" (*R* 6:32). Kant's reconciliation of these claims appeals to his Critical concept of freedom, such that free choice explains one's empirically-observable nature. More particularly, when Kant explains why character is rare, he shows how rarity is due to moral failing – reliance on inclinations – for which individuals are rightly held accountable (*VA*, 7:294). Rachana Kamtekar articulates a similar point to explain situational variation with respect to deception:

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 (of deception or non-deception) across situations . . .

[T]he absence of a strong enough interest . . . may help to explain cross-situational inconsistency.

(Kamtekar "Situationism, 269-70)

Situationists emphasize, in particular cases, that "the deeds in question do not require heroic commitment or sacrifice" (Doris *Lack*, 31). But Kant and Kamtekar point out that *developing a character* that acts consistently may require substantial (and thus rarely undertaken) sacrifices.

¹⁶ Barbara Herman "Making Room for Character," *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*, Ed. Stephen Engstrom and Jennifer Whiting (Cambridge University Press, 1996). Most of this essay is focused on Barbara Herman, *The Practice of Moral Judgment* (Harvard University Press, 1993). This paper was written before the appearance of Barbara Herman, *Moral Literacy* (Harvard University Press, 2007), so I do not address that work.

¹⁷ *Practice*, 78.

¹⁸ *Practice*, 75.

¹⁹ *Practice*, 91.

²⁰ *Practice*, 89.

²¹ *Practice*, 89, n14.

²² *Practice*, 78.

²³ *Practice*, 81, emphasis added, cf. 78.

²⁴ *Practice*, 43.

²⁵ *Practice*, 90.

²⁶ For further discussion, see Patrick Frierson, “The Moral Importance of Politeness in Kant's Anthropology,” *Kantian Review* 9 (2005): 105-27.

²⁷ *Practice*, 89.

²⁸ This point deserves particular emphasis given the many who criticize Kant for misunderstanding the complexities of maxim formation. Kant was aware not only that the same actions could be represented under different maxims, but also of the human tendency to use this fact as a means of justifying evil deeds.

²⁹ Throughout, I use Herman's summaries of Williams. Cf. Bernard Williams, *Problems of the Self* (Cambridge University Press, 1973), 207-29 and *Moral Luck* (Cambridge University Press, 1981), 1-19.

³⁰ *Practice*, 24.

³¹ *Practice*, 41.

³² *Practice*, 25.

³³ *Practice*, 31.

³⁴ *Practice*, 32.

³⁵ *Practice*, 34.

³⁶ *Practice*, 33.

³⁷ *Practice*, 35.

³⁸ *Practice*, 36-7.

³⁹ Herman “Character.”

⁴⁰ From Schiller's well-known and oft-quoted satire. See Allen Wood, *Kant's Ethical Thought* (Cambridge University Press, 1999), 23.

⁴¹ *Practice*, 24.

⁴² *Practice*, 37-8.

⁴³ *Practice*, 39.

⁴⁴ *Practice*, 40.

⁴⁵ *Practice*, 37-8.

⁴⁶ *Ibid.*

⁴⁷ Typically, Kant's idealism is interpreted to refer to either two-standpoints (see Henry Allison, *Kant's Theory of Freedom* (Cambridge University Press, 1991), Christine Korsgaard, *Creating the Kingdom of Ends* (Cambridge University Press, 1997)) or two worlds (see Eric Watkins, *Kant and the Metaphysics of Causality* (Cambridge University Press, 2005)). For discussion, see Frierson *Freedom* (chapter 1) and “Two Standpoints and the Problem of Moral Anthropology” (under review).

⁴⁸ *Lack*, 112.

⁴⁹ For discussion and further references, see Frierson, *Freedom*, 114-22.

⁵⁰ E.g. R 6:191; SF 7:43-4.

⁵¹ For further defense of this claim, see Patrick Frierson, “Review: Richard Dean, *Kant and the Value of Humanity*,” *Notre Dame Philosophical Reviews* 2007.04.17; Korsgaard *Creating*, and Wood *Ethical Thought*.

⁵² For elaborate discussion of this “humility,” see Jeanine Grenberg, *Kant and the Ethics of Humility* (Cambridge University Press, 2005).

⁵³ See Sharon Anderson-Gold, *Unnecessary Evil* (SUNY, 2001), Frierson *Freedom*, Wood, *Ethical Thought*, and Allen Wood, “Unsociable Sociability: The Anthropological Basis of Kantian Ethics.” *Philosophical Topics* 19 (1991): 325-351.