10

The Good, the Bad, and the Badass

ON THE DESCRIPTIVE ADEQUACY OF KANT'S CONCEPTION OF MORAL EVIL

PART 1 OF KANT'S four-part *Religion within the Boundaries of Mere Reason* is entitled, "Concerning the indwelling of the evil principle alongside of the good or Of the radical evil in human nature." In it, Kant is concerned with the age-old question of whether human beings by nature are morally good or morally evil. In his 1755 *Discourse on Origin and Grounds of Inequality*, Rousseau took the position that "Men are evil. Grim and constant experience dispenses us from the effort of providing a proof of this. I have however proven, as I believe, that man is good by nature."[1] In apparent direct opposition to Rousseau, Kant wrote: "If it is said, The human being is created good, this can only mean nothing more than: He has been created for the *good* and the original *predisposition* in him is good; the human being is not thereby good as such . . ." (R 6:43). Kant's position, then, was that despite being predisposed to moral goodness, human beings by nature are nevertheless evil. This evil is "radical" since it not only "corrupts the ground of all maxims," but "as a natural propensity, it is also not to be *extirpated*" (R 6:37). Kant develops his position in some detail in *Religion* 1, which, in addition to addressing the guiding question, represents an important advance in both his conception of moral accountability and moral psychology over his earlier moral writings in the 1780s.

1. The material in this chapter was given as the 2016 *Rousseau Lecture*, University of Keele, Keele, England, March 17, 2016, and as the annual *Parcells Lecture*, University of Connecticut, Storrs, CT, October 28, 2016. Rousseau's remark ~~is~~ from *Discourse on the Origin and Grounds of Inequality*, 1755, Part 1, Note IX.

Kant's conception of and argument for his radical evil hypothesis has attracted much critical attention. For example, critics argue that the view is unenlightening when it comes to explaining evil; that it involves a simplistic account of human nature; that it is ultimately incoherent in its claim that radical evil is something for which one is responsible despite being innate; and that Kant fails to argue adequately for the doctrine.[2]

Another criticism, and the one I address in this chapter, concerns the descriptive adequacy of Kant's theory of moral evil: that is, whether the view can accommodate pre-theoretical, commonsense views about evil.[3] I have in mind two related concerns. The first is about breadth: whether Kant's theory is capable of recognizing all of the various basic forms of moral evil that commonsense recognizes. Kant offers a three-fold taxonomy of basic types of evil, but the worry is that it misses important "middle-ground" forms of evil. Claudia Card raises this worry about breadth when she remarks that Kant "rejects the common-sense views that some of us lack basic commitments and that some of us have plural and incompatible but equally basic commitments" (2010: 87). This particular objection stems from Kant's thesis of character rigorism, according to which, by nature, individual human beings are either good or evil, they cannot be partly good and partly evil, nor can they be neither good nor evil.

A second, related worry concerns the so-called depth of Kant's theory, whether his theory of evil recognizes the varying magnitudes of evil manifested in human behavior that commonsense recognizes. For example, when discussing moral depravity in the *Religion*—the highest grade of evil for Kant—there is no mention of differences in the magnitude of evil involved, say, in lying about one's age on a job application compared to the evil manifested in brutal acts that result in great harm. As Todd Calder observes: "It seems far worse to torture someone for sadistic pleasure than to tell the truth to gain a good reputation. In fact, it seems reasonable to suppose that the first act (sadistic torture) indicates an evil will while the second act (telling the truth for self-interest) indicates a will lacking in moral goodness. But, for Kant, both acts indicate wills that are equally evil" (2015: 15). This concern with the depths of evil is brought into relief by the fact that there seem to be individuals who are "badasses," to use an apt label employed by Claudia Card in some of her work on moral evil.[4] Portrayal of such characters populate

2. For a useful presentation of and reply to these objections, see Louden 2010.

3. This objection relates to complaints that Kant's theory of human nature is overly simplistic. Therefore, in addressing the worry about descriptive adequacy, I will be implicitly addressing the issue of over-simplicity.

4. See Card 2017, who refers to the work of Jack Katz, in particular to chapter 3, "The Badass" of his 1988 book.

the world of crime fiction—often inspired, unfortunately, by real individuals. However, Kant explicitly denies the human possibility of a "diabolical will," someone who does evil for evil's sake—someone who apparently fits Card's description of a badass. The concern about depth, then, is whether and how Kant's theory of moral evil can make sense of variation in degrees of moral evil, both in deed and in character.

I agree with these critics that Kant's texts, particularly the *Religion*, apparently leave him open to these problems of descriptive adequacy. My principal aim in this chapter is to develop an interpretation of Kant's position on evil that I argue fits the texts reasonably well and avoids the objections in question. If my reading is successful, then Kant's psychology of moral evil is arguably more plausible than his critics have supposed.

To lay my cards on the table, my proposed interpretation (or reinterpretation) for dealing with matters of descriptive adequacy involves the following five claims. *First*, I understand the adequacy objections as directed against Kant's *empirical* moral psychology and not his a priori *transcendental* moral psychology. *Second*, I propose a reading of Kant's character rigorism that restricts it to the realm of his transcendental moral psychology, and therefore does not apply to his empirical moral psychology. *Third,* rigorism so restricted helps Kant's empirical moral psychology address middle-ground cases. *Fourth*, Kant's conception of the vices enables his view to deal adequately with concerns about the depths of moral evil. *Fifth*, properly interpreted, Kant's denial of the possibility of a diabolical human will is compatible with recognizing the moral evil characteristic of the badass.

In developing this line of interpretation, I will proceed as follows. Section 10.1 provides a brief overview of *Religion* 1 for purposes of orientation. In Section 10.2, I consider Kant's rigorism, making a case for its restriction. In Section 10.3, I turn to Kant's conception of moral goodness, as background for elaborating in Section 10.4 an interpretation of Kant's three grades of evil (frailty, impurity, and depravity). I argue that accommodating cases of impurity requires that one recognize mere lack of good will as a generic type of character flaw. Mere lack of good will falls between having a good will and having an evil will, and allows me to explain in Section 10.5 how recognition of this type of flaw helps somewhat to address the middle ground cases. Section 10.6 addresses concerns about matters of depth, and Section 10.7 considers the case of the badass. Finally, in Section 10.8, I conclude with some very brief remarks about the explanatory plausibility of Kant's conception of evil in light of contemporary work in social psychology.

10.1. Major Themes in *Religion* 1

In order to provide a proper setting for what is to follow, let us first briefly review some of the major themes from Part 1 of *Religion*.[5] *First*, within the context of the *Religion*, Kant's aim in Part 1 is to propose a secular interpretation of the Christian doctrine of original sin, one he develops from within his Critical philosophy. *Second*, in developing his secular interpretation, Kant traces the source of moral evil to a fundamental characteristic of the human being's power of choice (*Willkür*) that represents a particular orientation of this power in relation to the moral law (to be taken up in more detail in the next section). *Third,* Kant's concern is not with evil actions (actions that violate the letter of the moral law) but with evil maxims. "We call a human being evil, however, not because he performs actions that are evil (contrary to law), but because these are so constituted that they allow the inference of evil maxims in him" (R 6:20). More generally, his concern is with the fundamental ground of particular evil maxims and thus with matters of character.

Fourth, Kant is a normative-motivational dualist. Human beings have an original predisposition to "personality" or moral goodness, which (in part) involves the fact that such beings are aware of the moral law that grounds reasons for action and attitude that are normatively authoritative for such beings and who can be moved by such reasons to act accordingly. However, human beings are creatures with a sensible nature, and thereby have interests pertaining to their own happiness that provide reasons of self-love for action and attitude that can come into conflict with reasons grounded in the moral law. *Fifth,* moral evil involves allowing reasons of self-love to trump moral reasons. Kant refers to the source of evil as a "*propensity* to evil," roughly the innate tendency, characteristic of human beings' power of choice, to allow reasons of self-love to trump moral reasons.[6] *Sixth,* although this propensity is innate, paradoxically it is something for which one is morally accountable. *Seventh,* Kant's explanation of this accountability is in terms of a timeless choice one makes as a noumenal being, a choice which itself is inscrutable to human beings, given limits on human knowledge. Because the propensity to evil, itself an evil (R 6:32), is something for which one is accountable, and moreover "entwined with humanity itself, and as it were, rooted in it . . . [we] can further call it a *radical* innate *evil* in human nature" (R 6:32). *Eighth*, although this propensity is inextirpable, one's fundamental moral mission in life is to overcome it by striving to become a person of virtue thereby realizing as fully as one can one's original predisposition to moral goodness.

5. Later sections elaborate some of these themes.

6. This rough characterization is formulated moral precisely in the following section.

Finally, another claim that plays an important role in Kant's defense of the thesis of radical evil, and one of particular concern in this chapter, is his understanding of the controversy under consideration in Part 1. He explains:

> At the basis of the conflict between the two hypotheses . . . lies a disjunctive proposition: *The human being is* (by nature) *either morally good or morally evil*. It will readily occur to anyone to ask, however, whether this disjunction is accurate; and whether some might not claim that the human being is by nature neither of the two, others that he is both at once, that is, good in some parts and evil in others. Experience even seems to confirm this middle position between the two extremes. (R 6:22)

In response to the possibility of there being these middle position options, Kant first comments, "It is of great consequence to ethics in general, however, to preclude ~~as~~ far as possible, anything morally ~~indeterminate~~, either in actions . . . or in human characters; for with any such ambiguity all maxims run the risk of losing their ~~determinacy~~ and stability" (R 6:22). He then proceeds to argue on a priori grounds that the middle positions are not actually genuine options.

As I mentioned at the outset, it is Kant's rigorism, as typically interpreted, that is the basis for objecting that his conception of moral evil fails to accommodate what commonsense regards as genuine types of character flaw. It is with Kant's rigorism that I want to begin my defense against this particular objection.

10.2. Rigorism Restricted

In order to clear a path for accommodating middle-ground cases, I am proposing (on Kant's behalf) that his rigorism be understood as applying *exclusively* to options confronting human beings as members of the "intelligible" world, confronted with making the timeless choice mentioned previously. This is to deny that it applies to human beings as embodied members of the temporal "sensible," empirical world. As noted at the outset, to do so is to construe the thesis as pertaining to Kant's a priori transcendental moral psychology, but not his empirical moral psychology.[7]

In order to develop this idea, we first need to draw a distinction, making use of Kant's notions. It is perhaps tempting to suppose that "*böse Gesinnung*"

7. On the distinction between transcendental and empirical psychology in Kant's work, see Frierson 2014: 43–51, who summarizes the distinction as follows. "Transcendental psychology is a priori, offered from within the perspectives of our actively employed faculties, and normative. By contrast, empirical psychology is empirical, based on *observation* (even if in inner sense), and descriptive" (see p. 45). There is one passage in the Paralogisms chapter in

("evil disposition") and "*Hang zum Bösen*" ("propensity to evil") express equivalent notions. However, as Pablo Muchnik has pointed out, this temptation should be resisted.[8] As Muchnik explains, it cannot be Kant's view that these notions (or the terms that express them) mean the same.

> Otherwise, our personal wrongdoing would be explicated (and exculpated) by sheer membership in humanity []. This untoward conclusion, however, can be averted once we realize that the notions in question refer to two different units of moral analysis: the *Gesinnung* indicates the fundamental moral outlook of an *individual agent*, the propensity, the moral character imputable to the whole species. Overlooking the logical independence of these analytic units gives the impression that Kant's talk of a universal propensity to evil is inconsistent with his commitment to freedom. For if we consider "*Gesinnung*" and "propensity" to be synonymous, it seems natural to suppose that the choice at the level of the species carries causal efficacy at the level of the individual, and hence is at odds with our autonomy[]." (Muchnik 2010: 117, the empty brackets indicate deleted footnote numbers)

A slightly different way to make Muchnik's point is that if one conflates the notions in question, then it would not be possible to overcome one's propensity to evil; one would be stuck with an evil *Gesinnung*. It is Kant's view that individuals, as members of the species, have an inextirpable propensity to evil, yet as embodied agents they are able to overcome this propensity by coming to be virtuous. Following Muchnik, then, I will understand Kant's use of "propensity" in the present context as referring to the moral character of the species, and *Gesinnung* (or "disposition" [9]) as referring the moral character

---

the A edition of the *Critique of Pure Reason* where Kant refers to transcendental psychology, cosmology, and theology as "putative sciences of pure reason" (A 397), thus dismissing them as not genuine sciences. In this passage, "transcendental psychology" refers to an illusory science of the soul, understood as a persisting immaterial substance. Kant typically refers to this illusory science as "rational psychology," which the Paralogisms chapter attempts to undermine. However, as Frierson is using the label, transcendental psychology does not purport to theorize about the nature of an immaterial soul, and so is not illegitimate by Kant's lights. (I thank Houston Smit for discussion of this matter.)

8. Muchnik cites a passage in Allison 1990: 153 in which the two notions are equated.

9. The English translation of *Gesinnung* ~~is~~ simply "disposition," which we find in the Cambridge edition of Kant's *Religion* is problematic. On this point, see Munzel 1999: xvii–xviii, who prefers "comportment of mind" that is indicative of the sort of principled mindset, which may be good or evil, as a translation that captures the sense Kant assigns to *Gesinnung*.

of individual members of the species. That Kant's text seems to demand this distinction is reason to accept it, independently of my use of it.

Since I am proposing to restrict Kant's rigorism to the options available for timelessly choosing the character of one's power of choice, let us consider the key elements that figure in this choice. *First,* Kant represents this free timeless choice as resulting in the adoption of a maxim, which he refers to as the "subjective ground . . . of the exercise of the human being's freedom in general (under objective moral laws) antecedent to every deed that falls within the scope of the senses" (R 6:21).[10] *Second,* as noted in the previous section, what this maxim represents is a fundamental orientation of the faculty of choice in relation to the moral law. *Third,* Kant claims that this first subjective ground "can only be a single one, and it applies to the entire use of freedom universally" (~~ibid.~~). Presumably, what Kant means by "single" is just that there is a single fundamental maxim that constitutes the propensity in question. By "universal," he is apparently referring to the fact that either the propensity characteristic of one's power of choice is wholly good, or it is wholly evil.[11] *Fourth,* although Kant expresses this timeless choice as adopting one or another global maxim, the choice in question concerns the motivational *structure* of one's power of choice. *Fifth,* this choice is best understood

10. With respect to the maxim in question, which for Kant ultimately underlies the adoption of evil maxims by embodied human beings, he remarks, "One cannot, however, go on asking ~~which~~, in a human being, might be the subjective ground of the adoption of this maxim rather than its opposite" (R 6:21), otherwise it cannot count as an exercise of freedom. In this quote, Kant is distinguishing the maxim one adopts via a timeless choice from the *ground* or the reason for the adoption of the maxim. What is fundamentally inscrutable is the reason or ground that would explain why one chooses as one does. Kant's text would have been clearer had he distinguished the maxim in question from the "first subjective ground" of the choice of this maxim. Alternatively, if the maxim in question includes reference to one's grounds, then it would have been better had he clarified that it is the grounding factor that is the inscrutable first ground. In fact, in a later passage, Kant hints at this very thing. Speaking of the limits of explaining the ultimate source of evil, he says, "~~W~~e cannot derive this disposition, *or rather its highest ground,* from a first act of the power of choice in time. . ." (R 6:25, my emphasis). Reference here to disposition is really, given the proposed regimentation of terminology, reference to the propensity characteristic of the power of choice. Thus, it is not disposition (understood as a maxim) that cannot be derived, it is the basis or ground upon which one chooses the propensity in question that cannot be derived and is thus inscrutable to human beings.

11. The key passage for interpreting the two notions in question is at 6:24–5, where Kant is arguing against the "syncretist" position that the human being by nature is in some parts good and in some parts bad. The passage reads: "For if he is good in one part, 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."

(I claim) as choosing a will (*Willkür*) that has one or another general motivational tendency or "bent."[12] *Sixth*, in light of the third point, and given Kant's normative-motivational dualism and the fact that reasons of self-love can come into conflict with moral reasons, what one is choosing is to have a will with one of the following motivational tendencies:

> GOOD: The tendency not to allow violations of the moral law on behalf of self-love by subordinating reasons of self-love to moral reasons.
> EVIL: The tendency to allow (at least occasional) principled violations of the moral law on behalf of self- love, by subordinating moral reasons to reasons of self-love.[13]

Seventh, for reasons inscrutable to human beings, one chooses Evil, which, according to Kant is a fundamental characteristic of the species as free agents.[14] *Eighth,* in discussing the propensity to evil, Kant makes it clear that the "unit" of analysis (to borrow from Muchnik) is the species. He remarks that in considering whether the human being is good or evil "we are entitled to understand not individuals (for otherwise one human being could be assumed to be good, and another evil by nature) but the whole species" (R 6:25). However, this is not to say that individuals may come to lack this propensity; they never do as members of the species. *Ninth*, although the propensity to evil is inextirpable, it can be overcome. It can be overcome by one's coming to have a good disposition; by coming to have and exercise a good will. I understand this to mean that by coming to have and exercise a good will, one's inextirpable

12. Obviously, the idea of choosing one's fundamental orientation of one's power of choice is paradoxical to say the least. For a brief mention of how Kant attempts to handle this apparent paradox, see n. ~~19~~.

13. I understand Kant to identify moral evil with depravity—a principled violation of the moral law. My formulation here is intended to reflect this. However, this tendency is also manifested in cases of frailty and impurity, which Kant identifies as "grades" of the propensity to evil. He claims that the "origin" of depravity is "frailty of human nature, . . ." coupled with "dishonesty in not screening incentives" (R 6:37) resulting in impurity. See also n. 26.

14. Here is an appropriate place to mention a possible source of confusion, related to the distinction between propensity and disposition, namely, Kant's reference to a "supreme maxim." Given Kant's position on the controversy over human nature, in the context of discussing the propensity to evil, reference to one's supreme maxim refers to the maxim associated with Evil. However, in the context of considering embodied individuals, any reference to one's supreme maxim should be taken to refer to the particular moral character of the individual (her or his *Gesinnung*), which may or may not be evil—it is possible to overcome the propensity to evil by becoming morally good, even though the propensity to evil is "inextirpable". (Further, as I will argue on Kant's behalf, it is possible to lack a supreme maxim, so understood, as an embodied individual.)

propensity to evil is "masked," to use terminology from contemporary discussions of dispositions.[15] One's propensity to evil is masked when in circumstances where otherwise one's propensity to subordinate morality to self-love would be effective, its efficacy is blocked by one's good disposition.

Note that distinguishing propensity from disposition in the way suggested does not *entail* that the character (*Gesinnung*) of particular embodied individuals falls in the middle, being neither (wholly) good nor (wholly) evil or being a mixture of both good and evil. Rather, my proposal to restrict Kant's rigorism merely opens up the possibility that embodied individuals may have an empirical character that falls in the middle. The case for affirming middle-ground cases is in Sections 10.4 and 10.5.

Before moving forward let me comment briefly about my proposal and provide some reasons in its favor. First, I understand it to comport fairly well with the text of the *Religion* and other of Kant's texts. However, I am making a proposal on Kant's behalf that may not be what he intended; my project is to make Kant's psychology of moral evil as plausible as possible while preserving as well as possible key doctrines in his ethics. To re-conceive Kant's view of moral evil by proposing a wholesale rejection of rigorism (and all that would entail) would be a radical revision. Rigorism restricted preserves what I understand to be the essential core of the thesis and so strikes me as a relatively mild revision (if revision at all).

In defense of the restriction, note first that *Religion* 1 is primarily a work within Kant's a priori transcendental psychology. Kant remarks that the judgment to be made about the question of whether by nature human beings are good or evil is to be made according to "the scales of pure reason" and not from an empirical perspective that refers to embodied human individuals (R 6:25). Note, too, a footnote where Kant is commenting on his thesis of radical evil.

> [E]xperience can never expose the root of evil in the supreme maxim of a free power of choice in relation to the law, for, as *intelligible* deed, the maxim precedes all experience.—From this, i.e., from the unity of the supreme maxim under the unity of the law to which it relates, we can also see why the principle of exclusion [rigorism, M.T.] of a mean between good and evil must be the basis of the intellectual judgment of mankind, whereas, for empirical judgment, the principle can be laid down on the basis of *sensible deed[s]* (actual doing or not doing) that there is a mean between these extremes—on the one side, a negative mean of

15. See, for example, Johnston 1992.

> indifference prior to all education; on the other, a positive mean, of mixture of being partly good and partly evil. This second judgment, however, concerns only the human morality as appearance and in a final judgment must be subordinated to the first. (R 6:39n)

What the body of this passage apparently says contradicts the objection that Kant's empirical moral psychology rules out the middle-ground cases of indeterminacy and of mixture. I read the last sentence as saying that when it comes to judging "the *intelligible* deed" that precedes all experience, an appeal to empirical considerations must give way to the judgment demanded by reason. This sentence does not impugn judgments of character based on empirical evidence.

Finally, as Patrick Frierson explains in his book on Kant's empirical psychology, Kant's psychological theorizing at the empirical level is constrained by the competing aims of, on one hand, providing unified explanations of the empirical psychological facts, yet on the other, accommodating (without distortion) the facts to be explained.[16] At the empirical level of individual psychology, then, one would expect Kant to recognize a full range of types of character flaw, including middle-ground cases. In Sections 10.4 and 10.5 when I turn to Kant's conception of moral evil, this is what I shall argue. However, to set the stage for my argument, we must first consider Kant's theory of moral goodness.

10.3. Moral Goodness

It is part of Kant's transcendental psychology that all human beings have a predisposition to personality, that is, to moral goodness.[17] This predisposition includes having a moral conscience through which one is aware of the moral law and the reasons it grounds. According to Kant's normative moral theory, one's highest vocation in life is to fulfill the duty of self-perfection, which includes cultivating one's "natural powers" as well as one's predisposition

16. See Frierson 2014: 4–9.

17. As Kant defines "predispositions" (*Anlagen*), the term refers to those "constituent parts" of a being, as well as forms of their combination that "make for such a being" (R 6:28). They are "original" if they belong to the possibility of such a being by necessity; otherwise, they are merely contingent. With regard to human beings, the predispositions of interest pertain to one's power of choice and, viewed teleologically, they are tendencies that direct human beings toward certain good ends. Kant identifies three original (and thus necessary) predispositions to good that correspond to his tri-fold division of human nature into one's animality as a living being, humanity as a rational being, and personality as a morally accountable being.

to moral goodness. The natural powers include one's mental and physical capacities, "the highest of which is *understanding*, the faculty of concepts and so too of those concepts that have to do with duty. At the same time this duty includes the cultivation of one's *will* (moral cast of mind), so as to satisfy all the requirements of duty" (MS 6:387). In particular, a "human being has a duty to carry the cultivation of his *will* up to the purest virtuous disposition, in which the law becomes also the incentive to his actions that conform with duty and he obeys the law from duty" (MS 6:387). Virtue (pure virtuous disposition), then, is the full realization of one's predisposition to moral goodness and, as I propose to understand it, involves those elements mentioned or implied in the just-quoted remarks—namely, moral understanding, cultivation of a moral cast of mind, and strength. Here, briefly, is how I understand them and how they are related.[18]

Moral understanding fundamentally involves understanding the concept of duty that no doubt comes in degrees and involves some grasp of both the structure and content of what I will refer to as the normative moral realm. The content of this realm (for Kant) includes the basic ethical duties set forth in Part II of this work, "The Metaphysical First Principles of the Doctrine of Virtue." Thus, an essential element in striving to achieve one's moral vocation is to develop specific virtues that correspond to the basic positive duties in Kant's system while avoiding those vices associated with the basic negative duties. Part of having proper moral understanding, then, is to understand these virtues and vices and their importance in striving to achieve moral perfection. The structure of the normative moral realm involves taking moral reasons that impose strict requirements on one's behavior to be normatively

The predisposition to animality refers to basic animal drives directed toward the ends of self-preservation, propagation of the species, and community with other human beings. The predisposition to humanity refers to the natural tendency of a human being to use reason in the service of what Kant refers to as "self-love," whose ends include the happiness of individual as well as the advancement of civilization. The predisposition to personality is of main concern here. It is characterized as "the susceptibility to respect for the moral law *as of itself a sufficient incentive to the power of choice*" (R 6:27). This susceptibility for respect Kant refers to as moral feeling, and so the full realization of this predisposition results when one consistently acts on principles that incorporate moral feeling (so understood) into select maxims as one's sole and sufficient motive, specifically, into maxims whose immediate aim is to fulfill one's duties. This predisposition contrasts with the predisposition to humanity, in that it involves a use of reason that is not (or not just) in the service of self-love of the individual and the advancement of human culture. The end of this predisposition is human moral perfection (moral goodness), and represents that in virtue of which human beings are morally accountable.

18. For remarks the importance of moral understanding in Kant's conception of virtue, see VA 25:633.

superior to both reasons of self-love and to moral reasons that merely favor without strictly requiring action. It also involves taking moral reasons for action that favor without requiring (associated with the wide duty of beneficence) to provide one with sufficient (though not necessarily overriding) reasons for action in cases where such action would not violate a strict moral obligation or involve excessive self-sacrifice.

Of course, mere understanding of the basic content and structure of this complex normative realm is not sufficient for having a good will. In addition, one must also have a general commitment to comply with one's understanding—a "moral cast of mind." This is the "volitional core" of virtue. To have such a moral cast of mind includes having made a firm moral resolve to act only in ways that comply with the structure of the normative realm, including a resolve to cultivate one's cast of mind by avoiding vice and developing the virtues. Thus, when fully mature, the cast of mind in question involves the following complex resolution that constitutes what Kant refers to as a good will:

> *Moral resolve*. (a) In situations where one recognizes a perfect moral obligation to perform (or refrain from performing) an action (or series of actions), one resolves to comply by acting *solely* out of respect for the moral law, that is, because one recognizes that one has this kind of obligation. (b) Regarding general ends that one has a perfect moral reason to adopt (the obligatory ends of self-perfection and the happiness of others) one resolves to adopt such ends solely out of respect for the moral law, and then act to promote those ends on appropriate occasions. (c) Finally, one resolves to cultivate those particular qualities of character (the virtues) that dispose one to exercise good judgment in complying with such general obligations, while avoiding particular vices, again solely because this is part of what it is to be a virtuous person.[19]

With this conception of a fully mature good will in hand (composed of proper moral understanding and associated moral resolve), we can now characterize what it is to have a virtuous character; what it is that fully realizes one's predisposition to moral goodness. Virtue, in the highest degree possible for human beings, includes the following. (1) One coming to have the sort of "comportment of mind" (*Gesinnung*) just described, established through a firm resolve, and characteristic of a good will, which (2) is grounded in an understanding of the basic content and structure of the normative realm, together with

19. These include the character traits of beneficence, gratitude, and sympathetic feeling, which, together, compose the central duties of love toward others.

(3) the acquired strength of will to comply with one's resolve in particular circumstances.

Before moving forward, I conclude this section by adding to my characterization of moral resolve. To have genuine moral resolve involves a solemn vow to oneself whose content is what I have just explained. I understand the vow in question to include the following two constitutive characteristics. *First,* one's vow results in the resolve being *stable*, where stability refers to its unchanging persistence as an aspect of one's character once it is made. *Second*, one's vow must accurately reflect the *universality* of the resolve in the sense that it is taken to apply to all aspects of one's life, rather than to just a portion of it. I stress these two features of moral resolve because they play a significant role in the proposal I will make concerning the accommodation of those middle-ground cases that apparently have no place in Kant's conception of moral evil. *Finally,* in both the *Religion* and *Anthropology,* Kant pictures coming to have moral resolve as involving a "revolution in the mode of thought" in which, by a "single and unalterable decision a human being reverses the supreme ground of his maxims by which he was an evil human being" (R 6:47, see also A 7:294). This resolution is a matter of making or coming to have an unalterable firm resolve of the sort described earlier. Making this resolution is the first major step in striving to have a virtuous disposition; it is to have a good will. What one must then do is strive to acquire the strength of will to follow through on one's resolve. This is something that only happens over time during which one acquires those virtues that inoculate one from the vices. It is how one is able to overcome one's natural propensity to evil.

Having proposed a restriction on Kant's rigorism and sketched his theory of moral goodness (virtue), the tasks that remain are to consider whether and how Kant can accommodate commonsense judgments about moral evil by recognizing the breadth and varying depth of moral evil. In the following two section, we turn to matters of breadth.

10.4. Frailty, Impurity, Depravity

The propensity to evil is the principle mentioned in the title of Part 1 that exists "alongside" of the predisposition to good that characterizes the volitional nature of the human species.[20] In the previous section, I described this propensity as

20. Kant defines "propensity" (*Hang*) as "the subjective ground of the possibility of an inclination (habitual desire, *concupiscentia*), insofar as this possibility is contingent for humanity in general" (R 6:29). A footnote to this remark (added to the second edition) explains that a propensity is a kind of predisposition (*Prädisposition*) to acquire a desire for something antecedent to experiencing it, but when experienced "arouses" an inclination toward it. Kant's unsavory example concerns so-called savages who, he says, have a propensity for

the tendency to allow (at least occasional) principled violations of the moral law on behalf of self-love, by subordinating moral reasons to reasons of self-love. As a fundamental source of all particular instances of moral evil, explanations of all such evil trace back to this propensity. In Section II of Part 1, Kant introduces "three different grades of this natural propensity to evil"—frailty, impurity, and depravity—each representing a particular way in which the general propensity to evil can manifest in the character of embodied individuals. Here, I will propose an interpretation of these grades, arguing that accommodating them all requires recognition of a grade of the propensity to evil that falls between having a good will (*Gesinnung*) and having an evil will. This will open up room for accommodating the sorts of middle ground cases Card discusses.

However, before getting started, there is a matter of terminology to clear up. As I understand Kant's use of the term "*Böse*" ("evil") in *Religion* 1, he uses it in both a broad and a narrow sense. In the broad use, it refers to any flaw for which one is accountable, and so each of the grades of the propensity to evil are types of evil. Used in this way, it contrasts with moral goodness or virtue. Thus, for instance, the character flaw of frailty (weakness of will) is a species of evil, even though it is compatible with having a good will (see below). However, in various places in the text, Kant identifies moral evil with depravity, which I take to be moral evil in the true sense of the term intended by Kant (again, see below). I will be using the term in both its broad and narrow senses, making clear the sense of the term in use.

Let us proceed, then, to consider the three grades of the propensity to evil, which for present purposes involves two tasks. One task is to explain why it would make perfect sense for Kant to identify just the three grades of the propensity to evil he discusses in the *Religion*. A second task is to explain how Kant's conception of moral goodness nevertheless calls for the recognition of lack of moral resolve as a general type of moral failing that includes impurity, but which (as we shall see in Section 10.5) allows for middle ground cases.

The first task is not difficult. As explained in the previous section, for Kant virtue, or moral goodness of character, involves three essential elements: (1) having a good will (firm moral resolve) which involves a commitment to do one's duty, (2) from the sole motive of duty, and in addition (3) the strength to comply with the demands of duty. In recognizing the three particular grades of the propensity to evil in question, Kant is isolating types of character flaw that correspond to each of these elements. This explains why Kant considers just the three types of character flaw at issue. To bring this into focus, let us consider Kant's description of them.

The propensity to frailty (*Gebrechlichkeit*) isolates lack of strength and thus the tendency to fail to comply with the demands of the moral law. Here is Kant's description of this character flaw.

> [T]he frailty (*fragilitas*) of human nature is expressed even in the complaint of an Apostle: "What I would, that I do not!" i.e., I incorporate the good (the law) into the maxim of my power of choice; but this good, which is an irresistible incentive objectively or ideally (*in thesi*), is subjectively (*in hypothesi*) the weaker (in comparison with inclination) whenever the maxim is to be followed. (R 6:29)

One way to understand this passage is that the frail individual has made a firm resolve constitutive of a good will—to act solely out of respect for the moral law and thus solely on the basis of moral reasons in those circumstances in which one is to fulfill one's duties. However, when it comes time to comply with the letter of the law, the frail individual fails to do so, acting instead on a particular maxim whose motive is one of self-love. That Kant's conception of this malady is consistent with having a good will finds support in various passages. Consider, for instance, this one from the *Metaphysics of Morals*, where Kant is explaining that virtue requires governing affect (*Affekte*), that is, occurrent feeling states, such as sudden anger or joy that arise spontaneously and make rational reflection on choice and action difficult.[21]

> [A]n affect is called *precipitate* or *rash* (*animus praeceps*), and reason says, through the concept of virtue, that one should *get hold of oneself*. Yet this weakness in the use of one's understanding coupled with the strength of one's emotions is only a *lack of virtue* and, as it were, something childish and weak, which can indeed coexist with the best will. (MS 6:408)

---

into [illegible]ts and are thus predisposed to form "an almost inextinguishable desire" for alcohol once they have become acquainted with it. Although a person may have any number of particular propensities (some of them perhaps idiosyncratic), Kant is here interested in a global propensity—the "common ground" that applies universally to the human species—the propensity to evil, which he also refers to as the "~~first~~ subjective ground" of the exercise of one's freedom of the power of choice (R 6:2~~0~~). As first, it is "posited as the ground antecedent to every use of freedom given in experience (from the earliest youth as far back as birth) and is thus represented as present in the human being at the moment of birth—not that birth is its cause" (R 6:22, see also 6:42). To address this seemingly contradictory set of claims (an exercise of freedom antecedent to one's exercise of freedom), Kant distinguishes two senses of "deed"—the timeless choice of one's propensity is an intelligible deed, "cognizable through reason alone apart from any temporal condition," while the deeds performed as an embodied human being are "sensible, empirical, given in time" (R 6:31).

21. "Affect is surprise through sensation, by means of which the mind's composure . . . is suspended" (A 7:252).

I read this passage as implying that weakness of will is compatible with having a good will.[22] The weakness is not in one's moral understanding, nor, presumably, is it a matter of failing to have proper moral resolve; Kant says that this weakness can coexist with the best (i.e., good) will. The weakness is in not following through with one's moral resolve, a failure in the *use* of one's moral understanding.

If this is a proper reading, then one can have a general resolve to comply with duty, characteristic of a good will, but fail to follow through in particular circumstances because one's inclinations of self-love are stronger than are competing moral considerations, grounded in one's understanding.[23] Moral goodness is having a good will (firm moral resolve) plus the strength of will to comply with one's resolve. Even if an individual can suffer from moral frailty yet lack a firm moral resolve (which is surely the case), nevertheless, focusing on cases in which one has a good will, yet fails to comply with the moral law, serves to isolate *mere* lack of strength of will as type of character flaw.[24]

In discussing the three grades of the propensity to evil, Kant considers impurity to be a more serious flaw than frailty, but not as serious as depravity.

---

22. Unfortunately, Kant does not explain what he means by "the best will." He cannot of course, be referring to a virtuous disposition. However, the German here for 'will' is '*Wille*,' which, in its technical usage, refers to one's legislative capacity as a free agent and, in particular, the capacity to give oneself the moral law. It contrasts with '*Willkür*,' which, again in Kant's technical usage, refers to one's executive capacity as a free agent. I understand the goodness of a good will to refer to an individual's *Willkür*—one's executive power of free choice. In any case, it is incorrect to predicate good or evil of one's legislative capacity—a capacity that "cannot be called free or unfree" (MS 6:226).

23. Many interpreters have claimed that self-deception is essential to *akrasia*. In his illuminating 2015 "Irrationality and Self-Deception within Kant's Grades of Evil," Matthew Rukgaber argues that in contrast to impurity and depravity, *akrasia* does not result from self-deception. He goes on to distinguish impurity from depravity by the "level" of self-deception characteristic of each. He thus distinguishes the three grades in terms of distinct psychological mechanisms that produce each form of evil. My approach here is to distinguish them by relating each to a distinct aspect of Kant's conception of moral goodness without delving into details about the mechanisms responsible for the various propensities.

24. Well-known is the problem of understanding how weakness of will (*akrasia*) is possible. On the so-called standard view of *akrasia*, which takes as its model inter-personal deception, the akratic individual acts contrary to what, at the time of action, is her better *judgment* about what to do. According to an alternative revisionist view, *akrasia* often simply results from one's failure to act according to a previous *resolution* that one has not forsaken. Kant's reference to the complaint of the Apostle (in Romans 7:15) "What I would, that I do not!" suggests that Kant has the standard conception in mind. However, I see no reason that would prevent Kant from allowing that some cases fit the standard model, while others fit the revisionist model. For a helpful overview on the general topic of weakness of will see Stroud 2014. A related problem is how to understand akrasia given Kant's theory of action and the role maxims play within that theory. For ~~a~~ helpful discussion of this problem, and how it can be resolved, see Johnson 1998 and Frierson 2014: 232–48.

However, for the moment, I will skip over impurity and first consider depravity (*Bösartigkeit*). Kant writes:

> [T]he depravity (*vitiositas, pravitas*) or, if one prefers, the *corruption* (*corruptio*) of the human heart is the propensity of the power of choice to maxims that subordinate the incentives of the moral law to others (not moral ones). It can also be called the *perversity* (*perversitas*) of the human heart, for it reverses the ethical order as regards the incentives of a *free* power of choice; and although with this reversal there can still be legally good (*legale*) actions, yet the mind's attitude is thereby corrupted at its root (so far as the moral disposition is concerned), and hence the human being is designated as evil. (R 6:30)

As I understand this malady, it involves *resolving not* to always comply with one's intellectual apprehension of the structure of the normative realm by reversing the proper normative order of moral reasons and reasons of self-love. The evil consists in "subordinating" (R 6:36) moral reasons to reasons of self-love that thereby, according to Kant, constitutes a complete corruption of the mind's attitude "at its root." Such resolution is directly contrary to moral resolve, and this is Kant's conception of genuine moral evil.[25] As he remarks in a footnote at the outset of Part 2 of *Religion*, "genuine evil consists in our *will* **not** to resist the inclinations when they invite transgression, and this disposition is the really true enemy" (R 6:58, bold added). To relate this to the timeless noumenal choice of one's propensity discussed in Section 10.2, a depraved *Gesinnung* is the realization in a person's empirical character of the maxim, Evil.

Depravity, then, involves having a truly evil will (corrupt mind with respect to the normative realm). To isolate the essential element of subordination that is constitutive of an evil will, resulting in a will that is directly contrary to moral resolve, Kant describes a calculating prudentialist who complies with the letter of the moral law but only because he reasons that such compliance will promote his happiness.

---

25. Although frailty and impurity are character flaws, Kant to my knowledge never singles them out as species of true moral evil. He introduces them along with depravity as grades of the *propensity* to evil. They *hinder* the development of virtue and dispose one to become depraved. However, true moral evil for Kant involves more than hindrance, it involves the kind of principled *opposition* characteristic of depravity. Moreover, Kant concludes his description of depravity by remarking that such a human being is "designated as evil," something he does not say in his descriptions of the frail and impure.

> In this reversal of incentives through a human being's maxim contrary to the moral order, actions can still turn out to be as much in conformity to the law as if they had originated from true principles—as when reason uses the unity of the maxims in general, which is characteristic of the moral law, merely to introduce into incentives of inclination, under the name of *happiness*, a unity of maxims which they cannot otherwise have. (For example, when adopted as a principle, truthfulness spares us the anxiety of maintaining consistency in our lies and not being entangled in their serpentine coils.) (R 6:36–7)

Of course, many depraved individuals are not likely to make this calculation; it is more likely that such individuals will end up doing horrible things. However, again, Kant's aim is to isolate the fundamental source of moral evil in subordinating moral reasons to reasons of self-love, and the case of the calculating prudentialist does isolate this source.

Kant characterizes impurity (*Unlauterkeit*), the intermediate form of the propensity to evil, as follows.

> [1] although the maxim is good with respect to its object (the intended compliance with the law) and perhaps even powerful enough in practice, it is not purely moral, i.e., it has not, as it should be [the case], adopted the law *alone* as its *sufficient* incentive [2] but, on the contrary, often (and perhaps always) needs other incentives besides it in order to determine the power of choice for what duty requires; in other words, actions conforming to duty are not done purely out of duty. (R 6:30, bracketed numbers inserted)

In contrast to frailty, impurity need not involve a violation of the letter of the moral law. This much is clear. How it otherwise differs from both frailty and depravity has been a matter of scholarly dispute.[26] The remarks following the second inserted bracket make it tempting to interpret impurity as a form of frailty; however, where owing to good luck, one happens to have non-moral incentives that favor doing the morally right thing, and which serve to pick up any motivational slack that moral incentives alone happen to lack.[27]

---

26. Again, see Rukgaber 2015 who argues that in order for impurity to be a type of failing distinct from frailty and depravity, one should understand it to involve a particular form of self-deception. I agree with Rukgaber that the element of self-deception helps identify a kind of impurity distinct from the other two character flaws. He thinks that without self-deception putative cases of impurity turn out to be cases of depravity. However, for reasons I am about to explain, I do not agree.

27. This was my understanding of impurity in Timmons 1993.

While I do think that this is one kind of impurity, there is another kind, suggested by the first part of the passage. It should be enough for impurity if one adopts maxims to comply with her duties but routinely fails to adopt "the law alone as the sole and sufficient motive," even if the strength of her moral motivation does not need motivational assistance from motives of self-love. I have in mind someone who does not have a good will because she never vowed to make the moral law her *sole* motive for performing dutiful actions. Consider Mary who has no problem avoiding vice, owing to her temperament.[28] Therefore, she fulfills her perfect duties. Temperamentally, she is also a naturally sympathetic person and so performs meritorious acts of kindness for others. She recognizes that the fact that an action fulfills a duty is a sufficient reason to do it but routinely finds that fulfilling her duties, if not pleasant, is at least not unpleasant. She therefore routinely acts not from the sole motive of duty but from mixed motives. Mary's moral failing is that she lacks the proper kind of moral resolve; she has never concerned herself with vowing to act from the sole motive of duty in fulfilling her duties. Of course, were a case to come up where duty clashed with what Mary found pleasant or was in her self-interest, she would comply with duty and do so presumably out of the motive of duty. Mary does not suffer from frailty, but nor is she depraved. There is a difference, after all, between *not resolving* to act from the sole motive of duty and *resolving not* to. The latter is characteristic of depravity; the former fits someone like Mary.[29]

Now notice the following. If the sort of impurity we have been discussing (exemplified by Mary) results from a mere lack of moral resolve, then we have a category that fits between frailty and depravity, and so the second task mentioned earlier has been completed. How does recognizing that mere lack of moral resolve is a type of morally flawed character help with accommodating the middle ground cases? The answer is that it helps with cases of indeterminacy, but dealing with cases of fragmentation will take more work as we are about to see.

---

28. For Kant, temperament refers to one aspect of the character of human beings considered as merely products of nature and concerns fundamental attributes of an individual's sensibility and (for Kant) include the sanguine, the melancholy, the choleric and the phlegmatic. See A 7:286–91.

29. Here is a variant of the Mary case that makes a similar point. Suppose that Mary* believes that the motive of duty confers moral worth on dutiful actions, but that doing one's duty out of love, or sympathy, or compassion are also value-conferring motives, on a par with the motive of duty. Further, suppose that Mary* has vowed to comply with duty from what she takes to be morally significant value-conferring motives. According to Kant, Mary* has a false belief about proper moral motivation and thus a false conception of proper moral resolve. Mary* like Mary is not frail, nor is she depraved. Her "moral" motivation is not pure. She, of course, differs from Mary in having resolved to fulfill duties from motives other than duty.

10.5. Mischief in the Middle

As mentioned at the outset, Card criticizes Kant's conception of evil as not able to countenance so-called cases of moral indeterminacy and cases of moral fragmentation, which, she claims, commonsense recognizes as such. There are two issues here to be distinguished. The first is whether Kant's conception of evil can accommodate the cases in question. The second is whether, if so, Kant's handling of these cases agrees with the commonsense conception of them. With regard to the case of indeterminacy, I think the answer to both questions is "yes." With regard to moral fragmentation, I think the answer to the first question is "yes," and to the second question no. However, I do not think the "no" answer is particularly damaging to the overall adequacy of Kant's view.

Consider first Card's case of indeterminacy featuring someone who is morally capricious—a moral flip-flopper:

> Consider someone who is unpredictably irresponsible. Some days, she feels like not getting up for work (or like getting up and playing hooky) and so calls in sick, not from weakness but because then inclination just seems more important. Other days, she is moved by obligation, despite feeling it would be a great relief to stay home and unplug the phone. She does the right thing then because that is what seems most important *then*. This woman appears ambivalent—not frail, not even committed to self-interest, but basically uncommitted.... The common-sense view is that she is immature, has not "got her act together," has not yet developed a fundamental commitment (and possibly never will). Yet we also tend to hold that against her. (2010: 87)

This is someone who, as Card says, "exhibits *unpredictably* different patterns in the *same* contexts, a fairly common case" (2010: 88). As Card notes, it would be implausible for Kant to respond by claiming that this person's basic commitment changes often. The sort of commitment characteristic of moral resolve must be something *stable* in that it persists over a stretch of time. Pretty clearly, the person Card describes simply lacks moral resolve. She is not frail, as Card notes, and she does not fit Kant's characterization of depravity, of having a fundamental commitment that subordinates moral reasons to reasons of self-love.[30] If we recognize on Kant's behalf that mere lack of moral

30. Of course, she does allow reasons of self-love to motivate her when she wrongfully plays hooky, but this is also true of the morally frail, and so this fact about the flip-flopper is not sufficient for being depraved.

resolve is a character flaw, then along with select cases of impurity, one should include cases of moral indeterminacy characteristic of the flip-flopper as a species of this particular kind of character flaw.

Although Kant does not include the case of indeterminacy among the basic propensities to evil, he does address such cases when discussing the notion of character. In his *Anthropology*, Kant contrasts a person of principle with persons who act guided by the "unstable condition of instinct" (A 7:294).

> [T]o have a character signifies that property of the will by which the subject binds himself to definite practical principles [i.e., stable maxims, M.T.[31]] that he has prescribed to himself irrevocably by his own reason. Although these principles may sometimes indeed be false and incorrect, nevertheless, the formal element of the will in general, to act according to firm principles (not to fly off hither and yon, like a swarm of gnats), has something precious and admirable about it; for it is also something rare. (A 7:292)[32]

A person with a good will is a person of moral principle. A person with a depraved will is also principled, though mistaken in their evaluative priorities. Card's flip-flopper is someone who lacks character, so understood. Certainly, for Kant, lack of character is a serious character flaw.[33] I conclude that Kant's conception of moral evil does recognize this case of indeterminacy and that it does agree with commonsense that such cases are ones of indeterminacy.

The more complicated case is moral fragmentation in which common sense would allegedly judge this person partly good and partly evil. While

---

31. The reference to principles here must be to maxims rather than moral laws in order to accommodate instances of evil character. On this point, see Frierson 2006, who cites a passage in the anthropology lecture notes (25:1384–5) which has Kant making this explicit in lecture.

32. Kant describes the temperament of poets in a way similar to Card's unpredictable hooky player. He speculates about the seeming caprice of poets who by temperament differ from "lawyers and others in the learned professions" in having a "peculiarity, which concerns *character*, namely of *having no character*, but being capricious, moody and (without malice) unreliable" (A 7:249).

33. Kant's remarks about persons of principle, suggests that even someone like the early Roman dictator Sulla (c. 138 BC–78 BC), who Kant describes as having an evil will, and thus a depraved disposition, is less bad than someone lacking in principle. See A 7:293. This seems to conflict with Kant's ranking of depravity as morally worse than impurity. The only way I see to reconcile these claims is to interpret Kant as holding that depravity is at least instrumentally less bad than lack of character, because, as he writes, "By character [which he thinks Sulla does have, M.T.] one can get the upper hand over temperamental maliciousness" (ibid.). Kant thinks this is much harder to do, if one lacks commitment to principles.

the flip-flopper displays unpredictable reactions across the same contexts, the fragmented person displays a lack of unity of will in exhibiting systematically different and conflicting patterns of behavior in different contexts. Again, let us consider Card's example taken from the memoir of Sue William Silverman, daughter of Irwin Silverman who was chief counsel to the US secretary of the interior from 1933 to 1953.[34]

> [Silverman] played key roles in establishing statehood for Alaska and Hawaii, Philippine independence, the creating of the Puerto Rico Commonwealth, home rule for the Virgin Islands, Guam, and Samoa, and civilian rule of Japanese possessions after World War II. From 1954 to 1958 he was president of large banks. He was photographed with President Harry Truman, Adlai Stevenson, and other influential political figures. And he was a child molester. For many years he assaulted his daughter sexually, severely, locking her door at night in her bedroom, beginning when she was less than five. (2010: 89)

Here is Card's commentary on this case:

> Were those who placed this man in positions of public trust *totally* deceived about his character? Or did he have a good side and an evil side? He appears at first to have embodied the contradiction Kant thought impossible. If, however, we regard him as responsible for both patterns of behavior, and if they truly do manifest conflicting principles or priorities that he has, his character is not at its *most* basic level defined by these principles (hence, does not exhibit the contradiction Kant rejected). Rather, at its most basic level his character is defined by his failure to take responsibility for himself in a way that people with more coherent or conventional inclinations might never have to. This kind of failure is not captured by a formal maxim prioritizing self-interest. The task facing this man it to create a coherent self. Nor is his failure well captured by frailty. How much strength could it take not to rape one's five-year-old daughter and continue doing so behind a locked door for years? There is a policy here, not a lapse. (ibid.)

For this Jekyll/Hyde case to challenge Kant's apparent denial that a person could be a mixture of moral good and evil, we would have to interpret

34. S. W. Silverman, *Because I Remember Terror, Father, I Remember You* (Athens, GA: University of Georgia Press, 1996).

Silverman's work as chief counsel as properly morally motivated—fulfilling his duties from the sole motive of duty. Card remarks that perhaps Silverman's behavior in his job was motivated by self-love (reputation and money). However, she also points out that it is possible that Silverman "made moral decisions conscientiously on the job (asking seriously whether he could universalize the maxims of his actions)."

In considering this case, and whether it is a problem for Kant's conception of moral evil, we first have to ask ourselves whether common sense agrees with Card that a global evaluation of Silverman is that he is partly good and evil. I am not sure. My impression of Silverman was that he was a moral monster, ~~despite subordinating~~ self-interest to duty in his public life as chief counsel. However, leave this point aside. Let us instead consider how Card is conceiving of this case and why she thinks Kant's conception of moral evil cannot recognize it.

First, notice that Card is thinking that because Silverman lacks the kind of moral resolve that constitutes a good *Gesinnung*, he is not morally good. Recall that on the conception of moral resolve I propose on Kant's behalf, one's resolve must be both stable and *universal*. If Silverman did resolve to perform his public duties conscientiously, out of the sole motive of duty, then whatever resolve he had in carrying out his public duties, it did not extend to all aspects of his life; it lacked the essential element of universality. Therefore, Card is correct in claiming that on Kant's view, Silverman lacks a good *Gessinung*. Furthermore, according to Card's description of him, Silverman lacks an evil *Gesinnung* because having one would require that Silverman have a *global* commitment to subordinate morality to self-love, which he presumably did not have. Therefore, because Kant's conception of evil cannot recognize such cases of fragmentation, his view completely overlooks this sort of case, and so is at odds with common sense.

However, there is more to Kant's conception of evil than Card recognizes. Recall that evil is a matter of subordinating moral reasons to reasons of self-love. What Card overlooks is that this subordination may occur in two ways. The calculating prudentialist who embraces a global commitment to subordinate morality to self-love represents one way, which is how Card is thinking of moral evil on Kant's view. Yet another way is represented by someone who simply possesses some particular vice, which need not involve a global commitment to subordinating moral reasons to reasons of self-love. The particular vices that Kant discusses in the *Metaphysics of Morals* is an important element of Kant's theory of evil. Here is an appropriate place to bring them into view.

Evil is a matter of subordination, and on Kant's view of the particular vices, this is precisely what they involve. Kant connects vices with passions (*Leidenschaften*).

> A *passion* is a sensible *desire* that has become a lasting inclination (e.g., *hatred*, as opposed to anger). The calm with which one gives oneself up to it permits reflection and allows the mind to form principles upon it and so, if inclination lights upon something contrary to the law, to brood upon it, to get it rooted deeply, and so to take up what is evil (as something premeditated) into its maxim. And the evil is then *properly* evil, that is, a true *vice*. (MS 6:408)

The sort of principled opposition to the moral law that characterizes a particular vice involves subordinating certain moral reasons to reasons of self-love. This is what makes a vice properly evil. Suppose we agree with Card that Silverman lacks a global commitment to subordinating morality to self-love. It is still open to Kant to claim that Silverman is evil in light of the vice of malice that he apparently had toward his daughter. Furthermore, to suppose that the only way Kant can accommodate this case is to attribute to Silverman the kind of global commitment to subordinating morality to self-love, is based on a mistaken conception of the propensity to evil. Recall from Section 10.2, the normative orientation constitutive of this propensity is properly formulated as follows:

> EVIL: The tendency to allow (at least occasionally) principled violations of the moral law on behalf of self-love, because of subordinating moral reasons to reasons of self-love.

The parenthetical remark is important, indicating that this global tendency can be realized by having vices, regardless of whether one has some global commitment to subordinate morality to self-love. Therefore, I think Card is mistaken in claiming that Kant's conception of evil completely overlooks the case of Silverman.

I imagine that Card would respond by claiming that even if Kant's conception does not overlook this case, it still does not accommodate common sense because on that conception Silverman is partly good and partly evil, and according to Kant (on my reading) he is just evil. Yet I think Kant's view is not so far off from common sense as Card would have us believe. Let me explain.

I grant that on Kant's view, and speaking of a global assessment of Silverman's character, he was evil; he had an evil *Gesinnung*. However, at another level of description, it is also true that in certain parts of his life (we are assuming) Silverman displayed some morally admirable traits, while in other parts he displayed some morally deplorable traits. For this reason, he failed to have a unified self; he was morally fragmented. Certainly, Kant can say all this, which is a way

of partially accommodating what Card takes to be a commonsense reaction to Silverman.

Before summing up, I wish to raise an issue about Kant's conception of evil that the case of Silverman brings to light. The issue concerns the relation between particular vices and being an overall vicious person. A similar question concerns the relation between particular virtues and being an overall virtuous person: which virtues and how many are required for an individual to be an overall virtuous person.[35] If having any vice, for Kant, is sufficient for being an overall evil person, then it would seem that on Kant's view there is a very low bar for being evil, and this may go counter to commonsense. Because Silverman's consistent and cruel treatment of his daughter is horrifying, judging him an evil person is, I think, a fair verdict about his character. However, consider the vice of gluttony, which Kant discusses, and imagine a committed glutton. If this were his only vice, would we say that he is an evil person? Of course not! However, if any subordination of morality to self-love reveals an overall evil character, then Kant's view, after all, would not accommodate a range of commonsense moral judgments about character. Now notice that if Kant's conception of evil sets a very low bar for having an evil *Gesinnung*, the result is a conception of good and evil that verges on a kind of rigorism at the level of empirical character. Whether this is problematic for Kant's view, I will here leave open.

To sum up the last two sections, in Section 10.4 I argued that there is reason for Kant to recognize mere lack of good will (good *Gesinnung*) as a character flaw that occupies a middle ground between good and evil. I made a case for this by reflecting on certain cases of impurity. In this section, I argued that Kant's conception of evil does accommodate cases of moral indeterminacy of the sort Card describes: such individuals lack determinate character, which is a type of flaw. With regard to cases of moral fragmentation, I argued that at least in the case of Irving Silverman, Kant's view would be that he has an evil *Gesinnung*. Yet Kant's view can somewhat accommodate commonsense when one considers the particular traits—some admirable, some deplorable—Silverman displayed. Finally, the Silverman case raises the general issue about the relation between particular vices and one's overall character, and how in particular to conceive this relation on Kant's theory of moral evil.

Let us now move on to consider questions about the depth of evil on Kant's view.

---

35. For one treatment of this question, see Adams 2006: ch. 8.

10.6. The Depths of Evil

Recall Todd Calder's objection that Kant's conception of moral evil cannot discriminate degrees of evil and that truth-telling from the sole motive of self-love is as evil as sadistic torture for pleasure. This objection relies upon conceiving Kant's conception of moral evil entirely in terms of whether one has subordinated moral reasons to reasons of self-love. The thought, then, is that if two actions involve such subordination, then they are equally evil, contrary to commonsense judgments about a range of cases, including the ones Calder describes. Kant's emphasis in the *Religion* on overarching commitments and the subordination of moral reasons to reasons of self-love no doubt fuels this objection.

However, the resources for addressing this objection are not in the *Religion*. One must look to Kant's conception of the vices, in particular, the so-called vices of hatred that include envy, ingratitude, and malice. These vices correspond to the ethical duties of love (and the associated virtues)—beneficence, gratitude, and sympathetic feeling—that Kant discusses in the *Metaphysics of Morals*. In that work, his treatment of these duties and corresponding vices is relatively brief and he does not delve into the psychology of these vices. However, in the student lecture notes one finds remarks that clearly indicate that Kant's conception of moral evil does accommodate commonsense judgments about depth.[36]

In the Collins lecture notes, Kant is reported as saying:

> All three, ingratitude (ingratitudo qualificata) [aggravated ingratitude, hatred of a benefactor], envy, and *Schadenfreude*, are devilish vices because they evince an immediate inclination to evil. That man should have a mediate inclination to evil is human and natural; the miser, for example, would like to acquire everything; but he takes no pleasure in the other having nothing at all. There are vices, therefore, that are both evil directly and indirectly. These three are those that are directly evil. (VE 27:440)[37]

Here, the evil Kant is referring to concerns the evil of harming others. Vices in which one has an immediate inclination to harm others are what Kant calls "devilish" in the lectures. One has an immediate inclination (habitual desire) for something when one desires a state of affairs just for its own

36. What follows repeats some of the discussion of the devilish vices in Section 9.3.

37. The bracketed material is the footnote in the English translation edition to the Latin phrase.

sake. By contrast, the miser is moved by an immediate inclination to accumulate money in excess of his needs; this inclination does not include harming others. Note the mention of hatred. As Kant conceives these vices, particular manifestations of them may involve hate, and hatred adds to the degree of evil such traits realize. Consider this remark, again in the Collins notes, concerning the matter of degree.

> If this ingratitude increases so much that he cannot endure his benefactor and becomes his enemy, that is the devilish degree of vice, since it is utterly repugnant to human nature, to hate and prosecute those who have done one a kindness. (VE 27:439)

Being ungrateful toward one's benefactor is an evil—a vice. However, ingratitude accompanied by hatred represents, as mentioned in the passage, a "greater degree" of evil. Therefore, with respect to the evil manifested in one's possessing one of the devilish vices, Kant is clear that the evil realized can vary in degree.

Moreover, in setting forth his system of ethical duties, Kant sometimes compares vices with respect to the degree to which they manifest disrespect for the dignity of others and of oneself. Kant claims, for instance, that gluttony is worse than drunkenness—a greater evil (MS 6:427). Therefore, with respect to manifestations of a single type of vice and with respect to distinct types of vice, Kant's conception of moral evil recognizes variations in degree.

Let us go back to Calder's contrast between truth telling from self-interest, and sadistic torture. Truth telling from self-interest may or may not express an evil character. Kant does not condemn as evil complying with duty from non-moral motives. However, consider again the case of the calculating prudentialist who, according to Kant, has an evil character. It is such a case that Calder needs in order for his comparison to be apt. However, I do not see any reason why Kant cannot say that although both have an evil character, that nevertheless the sadist, because her behavior manifests the vices of hatred (let us assume), is far more vicious than the truth-telling prudentialist. Therefore, I do not think Kant's conception of evil fails to accommodate differences in degrees of moral evil.

10.7. The Badass

The "badass," according to Card, is someone who does evil things because they are evil—evil for evil's sake. The aspiring badass, she claims, is not someone who does evil things in order to be respected by other badasses. "A real badass ~~does~~

~~not~~ care what others think. To deserve their respect, one must not act for the sake of it. Rather, one must become a certain kind of person" (2016: 37). Card draws a comparison between the genuine badass and someone who has a Kantian good will. For Kant, in order to deserve happiness as well as to be worthy of the esteem of an impartial rational spectator, one must have a good will, motivated not by the desire to be esteemed or to be happy; rather one must be motivated by recognizing the independent worth of a good will. Only then is one worthy of the esteem of an impartial spectator and of the happiness that a good will merits. Similarly, in order to be a genuine badass worthy of esteem from other badasses, one must be, for example, cruel and ruthless, *and* become the kind of person who takes immediate satisfaction in exercising such qualities of character. This is Card's conception of someone who does evil for the sake of evil, is cruel for the sake of being cruel, ruthless for the sake of being ruthless, and so on.

A commonly voiced complaint about Kant's conception of evil is that he explicitly denies that human beings are capable of having what he calls a "diabolical will" (R 6:35). In commenting on the propensity to evil (depravity), Kant remarks that it should not be called "malice" because this term, taken strictly, would be "a disposition (a subjective *principle* of maxims) to incorporate evil *qua* evil for incentive~~s~~ into one's maxim (since this is *diabolical*)" (R 6:37). These remarks are taken as clear indication that Kant denies the possibility of a type of individual who seems not only possible but also actual, and thus as an indication that Kant's conception of human nature is too narrow, if not naïve.[38] Card writes:

> Kant's theory of radical evil in human nature is not radical enough to comprehend taking pride in being bad. Pride so grounded seems to presuppose what Kant called *diabolical* evil, doing evil for evil's sake, for which he found no basis in human nature (2016: 38).

If one supposes that Kant held this because he thought this putative type of evil is of a magnitude not possible for humans, then the possibility of the badass shows that Kant's view lacks breadth because it lacks depth.

Let us agree with Card that the badass she discusses is someone who counts as doing evil for evil's sake.[39] The question is whether Kant's view can

38. See, for example, Silber 1960 and Bernstein 2002.

39. I understand "doing evil for evil's sake" to refer to someone who, for example, takes satisfaction in the infliction suffering of others just because it is evil, and not because such a person as Stanley Benn puts it, "sees it, in some partial or distorted way, as a good, even for himself. He does not think of himself better off for it; he is no less disinterested in rejoicing in it than is a benevolent person who rejoices in someone's good fortune" (1985: 806). For

accommodate evil individuals of this sort without having to recant his claim about diabolical wills. I think it can once we address questions about the sort of diabolical evil that Kant denies is humanly possible and about the psychology of the badass.

Why, then, does Kant deny the possibility of a diabolical will? Notice first that in the quoted passage from the *Religion* two paragraphs ago, Kant is referring to a "disposition (a subjective *principle* of maxims)" and claims that one should not call this disposition "malice." However, malice is one of the principal vices of hatred, so Kant does not deny that human beings can do evil things motivated by malicious hatred. How, then, can Kant both deny that human beings cannot have a malicious disposition, yet recognize malice as vice?

Here is a proposal, informed by my earlier remarks about the need to distinguish claims that figure in Kant's transcendental psychology from those that belong to empirical psychology. Suppose we interpret (or re-interpret if you like) Kant's denial that depravity is malice, as referring to a claim about the nature of the human will and, in particular about fundamental sources of one's normative reasons for acting, rather than referring to one's empirical character (*Gesinnung*). Recall, that for Kant, human beings have a three-fold predisposition to good.[40] The two of interest here are the predisposition to humanity (the source of reasons of self-love) and the predisposition to personality (the source of the moral reasons, including the capacity to act for such reasons independently of reasons of self-love). Together these two predispositions constitute Kant's normative-motivational dualism of human nature. For Kant, there is no predisposition to evil, as such; rather, one's evil nature is a propensity to subordinate moral reasons to reasons of self-love. Indeed, Kant's view is that it is not possible to conceive of human beings as having a predisposition to evil. What would constitute, for Kant, such a predisposition?

To answer this question, one starts by imagining a being with a predisposition to evil instead of having a predisposition to the good, that is, by imagining a devil. In the Vigilantius lecture notes, we find a contrast between the evil in human nature and the evil nature of a devil. "It is this, . . . [the propensity to evil, M.T.], which may distinguish man from a devil, who views himself as governed by evil itself, and as author of the same, and who, therefore,

---

readers who resist the idea of doing evil to evil's sake, and so would not describe the badass in such terms, they may read this section simply as how Kant can accommodate the badass within his empirical psychology.

40. See n. 17.

without struggle or inducement, engages in no actions other than bad ones" (VE 27:572). In the Collins lecture notes, the idea of being who, by nature, is predisposed to "devilish evil" is the idea of a being "where there is no seed of good at all, not even a good will" (VE 27:317). These are claims about the fundamental nature of the types of beings in question, and therefore they are claims from the perspective of Kant's transcendental psychology.[41] Human beings are not devils. Rather, according to Kant, human beings occupy middle ground between angels and devils.

Perhaps, then, if human beings are not devils, they are part angel (having a predisposition to moral goodness) and part devil (a predisposition to malice). However, this is not possible, according to Kant. Given that human beings have a predisposition to moral goodness, to suppose they also have a predisposition to malicious evil (evil for evil's sake) involves a contradiction of the following sort. The predisposition to moral goodness, we have noted, is the source of moral reasons and associated motivation. Moral reasons include reasons to promote the happiness of others. To suppose that humans also have a predisposition to malicious evil is to suppose they have a source of reasons antithetical to moral reasons; that is, that by nature human beings have, for example, underivative reasons to pursue such ends as promoting the misery of human beings. For Kant, this would be to suppose that one's rational nature is inherently contradictory and so something impossible.[42] I believe this (or something like it) is the line of thinking involved where Kant is considering various alternative hypotheses for locating the ground of evil (R 6:35). One hypothesis is that the ground is one's sensuous nature and the natural inclinations originating from it. The other is that evil (or malice on the present assumption) is grounded in one's rational nature. Kant rejects the first because he thinks that such inclinations themselves bear no direct relation to evil and that we are not in any case responsible for them. He rejects the idea of "an evil reason" (i.e., reasons to pursue evil for its own sake) "because resistance to the law would itself be thereby elevated to incentive (for without any incentive the power of choice cannot be determined, and so the subject would be a diabolical being)" (R 6:35). On the present reading, then, raising resistance to the law as itself an incentive (i.e., as an underivative reason to act contrary to the law) is to conceive of practical reason as itself corrupt, something Kant thinks is not possible.

41. Notice that Kant's view is *not* that the very idea of a diabolical being is *conceptually* incoherent. He thinks that both devils and angels are possible beings.

42. On the kind of incoherence concerning directly conflicting normative reasons under consideration here, see Caswell 2007.

The proposal I am making, then, is to interpret Kant's remarks about diabolical evil as being about fundamental aspects of human nature—about the nature of the original predispositions that characterize human beings as rational, accountable agents. If we do this, and restrict Kant's denial of the possibility of diabolical human beings to a claim of transcendental psychology, it remains an open question whether it is possible for an embodied human being to become a badass as Card conceives them. Let us proceed, then, to consider this.

The prima facie case for supposing that Kant's empirical psychology recognizes badass psychology is, I think, clear. In the previous section, I called attention to Kant's characterization of the vices of hatred as involving an "immediate inclination" to evil and as thus directly evil, contrasting such directly evil vices as envy, ingratitude, Schadenfreude (and later in the *Metaphysics of Morals,* malice) with the vice of avarice that is only indirectly evil. A natural interpretation of having an "immediate inclination" to cause another person to suffer—to torture someone for the satisfaction it brings one to be doing evil—is that it is an inclination to do evil for evil's sake.[43] In the passage where Kant is contrasting vices that are directly evil with those that are only indirectly evil, he asks whether such immediacy "is human and natural" (VE 27:440). The passage continues:

> The question may be raised, whether the human soul contains an immediate inclination to evil, and thus a propensity for devilish vice. We call a thing devilish when the evil in man is carried to the point of exceeding the level of human nature, just as we call angelic the goodness that surpasses the nature of man . . . . There is reason to believe, however, that in the nature of man's soul there resides no immediate inclination to evil, but that its tendency ~~to~~ evil ~~is~~ only in an indirect fashion. (VE 27:440–1)

Kant goes on in this passage to claim that humans cannot be so ungrateful as to hate one's benefactor nor do they have an "immediate urge" to rejoice at another's misfortune. He concludes, remarking: "Man therefore has no direct inclination toward evil *qua* evil, but only an indirect one" (VE 27:~~440–1~~).[44]

---

43. Allen Wood 2010: 154 also makes this point.

44. However, the passage goes on to note that the germ of *Schadenfreude* seems to be apparent in young children, which one might use in arguing that human beings do, by nature, have an immediate inclination to evil. However, the passage overall seems to side with the claim that human beings have only an indirect inclination toward evil qua evil.

Given that Kant claims that one can be so ungrateful to one's benefactor as to hate him and thus come to have an immediate inclination to evil directed toward that benefactor, I think the passage just quoted is to be read as asking a question about the fundamental nature of the human being. At the most fundamental level of human nature, is it true of human beings that the "human soul" includes an immediate inclination to evil? A "no" answer to this question is compatible with allowing that embodied human beings can come to hate others to such a degree that one takes satisfaction in hurting them for no other reason than that to do so is to be cruel, the mark of a real badass.

To return to my proposal: Kant's denial of diabolical human beings is a denial from the perspective of a priori transcendental psychology. This is important because it allows Kant to maintain that human beings are not the kinds of creature who *by nature* have a predisposition to malicious evil, yet also allows that human beings are capable of becoming badasses and thus capable of doing evil for evil's sake.

I know of two considerations that stand in prima facie opposition to my claim that Kant's moral psychology accommodates the badass, so described. One is Kant's apparent acceptance of a version of the guise of the good—the idea that all intentional action is undertaken in the belief that the action or end aimed at is in some respect good. The other is that Kant's normative-motivational dualism rules out doing evil for the sake of evil. Let us consider these.

In the *Critique of Practical Reason*, Chapter II, "On the Concept of an Object of Pure Practical Reason," Kant considers "an old formula of the schools" (KpV 5:59) that *we desire nothing except under the form of the good, and nothing is avoided except under the form of the bad*. In commenting on this formula, Kant notes that the terms 'good' and 'evil' are ambiguous and can refer to two different concepts which in German are designated by *das Gute* (good) and *das Wohl* (well-being), and *das Böse* (evil) and *das Übel* (ill-being). Kant goes on to claim that it is doubtful that the formula is true if the concepts involved are those of well-being and ill-being, but "indubitably certain" if understood as saying that "we will nothing under the direction of reason except insofar as we hold it to be good or evil" (KpV 5:60). The key expression here is "under the direction of reason." The badass wills the gratuitous suffering he inflicts on his victims. He does not act under the direction of reason, according to Kant, because reason does not direct one to inflict gratuitous suffering. The badass does value inflicting such harm, but he need not suppose that such actions are *valuable* in the sense that he has objectively good reason to do what he does. If we understand Kant's version of the guise of the good in this manner, restricted to what one does under the direction of reason, then Kant's

guise of the good is not clearly at odds with recognizing the psychology of the badass. To defend this interpretation requires more clarification and defense than space allows, so I will leave the matter up in the air and file it under "to be taken up on another occasion" and move on to Kant's normative-motivational psychological dualism.

How is it possible, then, to become a badass, given Kant's normative-motivational dualism, and in light of the fact that human beings do not have a predisposition to malicious evil? Again, space does not allow a full treatment of the matter; however, in outline, I believe the story goes as follows.[45] First, as we know, according to this dualism, there are two fundamental sources of reasons for action—morality and self-love.[46] One is born with a propensity to elevate reasons of the latter sort over reasons of the former sort. This is the propensity to depravity. However, one can *become* a malicious person, taking immediate pleasure or satisfaction in the suffering of others. One can even go so far as to become a badass in the following way. One does not start life with a predisposition to do evil for the sake of evil, however, perhaps because one was abused as a child, one comes to be misanthropic—what Kant refers to as "an *enemy of humanity*" (MS 6:450). Indeed, one comes to hate humanity and hate it deeply, eventually coming to take immediate pleasure in making others suffer—a real badass. It is, after all, a common phenomenon that actions originally done merely as a means to some desirable end are later desired intrinsically. One attends a jazz concert merely in order to be with friends, having no desire to hear such music. Over time, after many such concerts, one comes to appreciate the nuance and subtlety of the jazz one hears and comes to enjoy it. Of course, even if this phenomenon is common, it is a further step to claim that Kant's

45. A full development of the sketch to follow would require addressing the difficult issue of Kant's conception of self-love including his psychological hedonism. I understand Kant's conception of self-love as *not* committed to some narrow form of psychological egoism. Kant, after all, refers to individuals who are naturally sympathetic, who "without any other motive of vanity or self-interest they find an inner satisfaction in spreading joy around them and can take delight in the satisfaction of others so far as it is their own work" (G 4:398). Sophisticated interpretations of Kant's psychological hedonism are advanced by Andrews Reath 1989 (reprinted in Reath 2006), and Barbara Herman 2000 (reprinted in Herman 2007). See also Richard McCarty 2009: 48–52. My basic claim here is that Kant's theory of non-moral action is consistent with the psychology of the badass.

46. Regarding self-love, in the second *Critique* Kant writes, "This propensity to make oneself as having subjective determining grounds of choice into objective determining grounds of the will in general can be called *self-love*" (KpV 5:74). Notice that this characterization of self-love does not commit Kant to egoism with respect to non-moral motivation; it leaves open what the objects of one's inclinations are.

empirical psychology can make sense of the aspiring badass who eventually succeeds in becoming one. What, then, might Kant say about the psychological mechanisms by which one becomes a badass? Indeed, is his conception of human psychology rich enough to address this question? Let us conclude this section by briefly considering these questions.

In "Taking Pride in Being Bad," Card considers the sort of process by which one could become a badass. In very rough outline, it goes as follows. She borrows Korsgaard's (1996) notion of a self-concept that allows for much variety in the sorts of self-concepts one can embrace. She then appeals to Lorna Smith Benjamin's (2005) incorporation of John Bowlby's (1969, 1989) Attachment Theory that Benjamin argues can illuminate perverse, irrational, and perhaps downright diabolical behavior. Finally, on the basis of these ideas, Card speculates that the badass is someone who, at some stage in life, becomes attached to someone else taken by the aspirant to be a badass, seeking to imitate the person's behavior and psyche—eventually taking pride in being a badass. With this story in mind, Card comments on Kant's moral psychology:

> Kant's position that evil in human beings is not diabolical now seems partly right and partly wrong. It seems right that there is no need to suppose a fundamental predisposition to the bad in human nature. But people can knowingly choose to do evil without believing it to be prudent, and it is possible to come to value being bad. A predisposition to form attachments to others, missing in Kant's moral psychology, could explain why some people come to take pride in being bad.... What seems most right about the Kantian denial of diabolical evil, from the point of view of attachment theory, is that attachment even to an immoral model is not *initially* diabolical, in Kant's sense of the term. (2017: 54)

I believe there is a place in Kant's moral psychology for something like attachment theory, even if it is not something that Kant considers. In the lectures on ethics, we find Kant theorizing about the source of the vices of hatred. He traces their origin to the *impulse* (*Trieb*) of emulation, implanted in human nature, whose purpose "really lay in inciting men to constant cultivation of greater perfection in themselves by comparison with others" (VE 27:678). However, this same impulse can lead to rivalry in which one works against the well-being and standing of others that results in "a side of human nature that has become malignant" (VE 27:678). Of course, in order to emulate someone, one must be able to imitate them—do what they

do, take on the attitudes they have. Moreover, imitation is something Kant mentions in a number of places, particularly in his work on pedagogy, where he stresses the importance of imitation in the proper formation of children. "Parents who are already educated are examples for imitation by means of which children form themselves" (Päd 9:447). Furthermore, commenting on the fact that concepts of the understanding, such as cause and virtue, though not drawn from experience, only arise on the occasion of experience, Kant remarks: "No human being would have the concept of virtue if he were always among rogues" (LM 28:233). If the rogues in question were badasses, then presumably being around only such individuals would result in one's becoming like them.

Although Kant does not provide a detailed psychological story about imitation, his views allow for an account of how, through attachment and imitation, human beings without a predisposition to malicious evil can come to value cruelty for the sake of cruelty: that is, they come to have an immediate inclination to harm others for the sake of harming them. Therefore, as far as Kant's normative-motivational dualism is concerned, I see no reason why Kant's empirical psychology cannot recognize the badass. If one counts the badass as someone who does evil for the sake of evil, then Kant's empirical psychology does countenance people who fit this description.

10.7. Conclusion

My concern throughout has been with the descriptive adequacy of Kant's conception of moral evil. I have made a case, based partly on restricting Kant's rigorism that his theory can accommodate the central types of case, including the badass, that Card and others appeal to in objecting to his conception of moral evil.[47] I have also made a case for the claim that, once Kant's conception of the vices of hatred are included in his conception of moral evil, his view can accommodate commonsense judgments about the varying magnitude of moral evils.

47. In addition to cases of indeterminacy, moral fragmentation, and the badass, Card 2002: 211–34 considers so-called grey zone cases exemplified by Auschwitz prisoners who accepted positions of ghetto police in charge of rounding up other prisoners to be sent to their death. She asks whether, despite being complicit in evils done to those prisoners, such individuals had an evil will. Explaining how Kant's moral psychology of evil would deal with such cases is a task for another article.

Of course, even if I have succeeding in doing all this, there remains the question of whether Kant's conception properly *explains* human evil. In *Religion* 1, the focus is on the ultimate source of evil being the corruption of the human will. Since the late 1960s, some social psychologists, and philosophers following them, have attacked the so-called dispositionalist explanation of behavior (which appeals to character traits in explaining evil) in favor of a situationist account. According to situationism, good people can be induced to perform evil deeds as a result of the situation in which they find themselves. In general, facts about one's situation rather than facts about character do the heavy lifting in explaining the evil people do. Situationism, then, viewed as alternative explanation of evil, seemingly represents a challenge Kant's apparent dispositionalist account.[48]

Here is not the place to get into this apparent challenge. I will end by just noting that Kant's primary aim in *Religion* 1 was to discover the fundamental *source* of moral evil, which he finds in the individual, as a member of the human species. Kant's hypothesis that human beings have an innate propensity to evil is compatible with claims about individuals lacking the sort of stable character traits featured in dispositionalist accounts, and it is also compatible (so I would argue) with situational forces triggering evil behavior. Regarding stable character traits, Kant remarks that "character is set very late, approximately by age forty, for one can there best separate the concepts from instincts" (HA 25:654). This is because at approximately that age, explains Kant, instincts and inclinations have lost their force, which allows one to settle on firm principles governing action. Furthermore, relying on empirical evidence, as Kant does (R 6:32–4), but without embracing Kant's metaphysical commitments, it is plausible that human beings have a deep-seated tendency to elevate self-love over morality. Situationism highlights just how frail even "good" individuals—individuals of good morals, but lacking in moral character—can be when they encounter challenging circumstances. Kant would agree.

48. The *locus classicus* for situationism is Walter Mischel's 1968 book. For a summary of key experimental evidence in favor of situationism, including details about Philip Zimbardo's Stanford Prison Experiment, see Zimbardo 2007. For remarks about the current state of situationism within social psychology and in philosophy, see Miller 2016. See also Wielenberg 2006, who defends Kant against the situationist challenge.

References

Adams, R. 2006. *A Theory of Virtue: Excellence in Being for the Good* (New York: Oxford University Press).

Allison, H. 1990. *Kant's Theory of Freedom* (Cambridge, UK: Cambridge University Press).

Benn, S. 1985. "Wickedness," *Ethics* 95: 795–810.

Bernstein, R. J. 2002. *Radical Evil: A Philosophical Interrogation* (Cambridge, MA: Polity Press): ~~36–42.~~

Bowlby, J. 1969. *Attachment and Loss: vol. 1, Attachment* (London: Hogarth).

_____. 1979. *The Making and Breaking of Affectional Bonds* (New York: Routledge).

Card, C. 2002. *The Atrocity Paradigm* (New York: Oxford University Press).

_____. 2010. "Kant's Moral Excluded Middle," in S. Anderson-Gold and P. Muchnik, eds., *Kant's Anatomy of Evil* (Cambridge, UK: Cambridge University Press): 74–92.

_____. 2016. "Taking Pride in Being Bad," *Oxford Studies in Normative Ethics* 6: 37–55.

Calder, T. 2015. "The Concept of Evil," *The Stanford Encyclopedia of Philosophy* (Fall Edition), Edward N. Zalta (ed.), http://plato.stanford.edu/archives/fall2015/entries/concept-evil/.

Caswell, M. 2007. "Kant on the Diabolical Will: A Neglected Alternative?" *Kantian Review* 12: 147–57.

Frierson, P. R. 2006. "Character and Evil in Kant's Moral Anthropology," *Journal of the History of Philosophy* 44: 623–34.

_____. 2014. *Kant's Empirical Psychology* (Cambridge, UK: Cambridge University Press).

Herman, B. 2000 [2006]. "Rethinking Kant's Hedonism," in *Moral Literacy* (Cambridge, MA: Harvard University Press): 176–202.

Johnson, R. 1998. "Weakness Incorporated," *History of Philosophy Quarterly* 15: 349–67.

Johnston, M. 1992. "How to Speak of the Colors," *Philosophical Studies* 68: 221–63.

Katz, J. 1988. *Seductions of Crime: Moral and Sensual Attractions in Doing Evil* (New York: Basic Books).

Korsgaard, C. 1996. *The Sources of Normativity* (Cambridge, UK: Cambridge University Press).

Louden, R. B. 2010. "Evil Everywhere: The Ordinariness of Kantian Radical Evil," in S. Anderson-Gold and P. Muchnik, eds., *Kant's Anatomy of Evil* (Cambridge, UK: Cambridge University Press): 93–115.

MaCarty, R. 2009. *Kant's Theory of Action* (Oxford and New York: Oxford University Press).

Miller, C. B. 2016. "Empirical Approaches to Character," *The Stanford Encyclopedia of Philosophy* (Fall 2016 Edition), Edward N. Zalta (ed.), http://plato.stanford.edu/archives/fall2016/entries/moral-character-empirical/.

Mischel, W. 1968. *Personality and Assessment* (New York: Wiley).

Muchnik, P. 2010. "An Alternative Proof of the Universal Propensity to Evil," in S. Anderson-Gold and P. Muchnik, eds., *Kant's Anatomy of Evil* (Cambridge, UK: Cambridge University Press): 116–43.

Munzel, G. F. 1999. *Kant's Conception of Character: The "Critical" Link of Morality, Anthropology, and Reflective Judgment* (Chicago and London: University of Chicago Press).

Reath, A. 1989 [2006]. "Hedonism, Heteronomy, and Kant's Principle of Happiness," in *Agency and Autonomy in Kant's Ethics: Selected Essays* (Oxford and New York: Oxford University Press): 33–66.

Rukgaber, M. 2015. "Irrationality and Self-Deception within Kant's Grades of Evil," *Kant-Studien* 106: 324–58.

Silber, J. 1960. "The Ethical Significance of Kant's *Religion*," in T. M Greene and H. H. Hudson, eds., *Religion with the Limits of Reason Alone* (New York: Harper & Row): cxxv–cxxvii.

Smith, L. B. 2005. "An Interpersonal Theory of Personality Disorders," in J. F. Clarkin and M. F. Lenzenweger, eds., *Major Theories of Personality Disorder*, 2nd ed. (New York: Guilford Press): 157–230.

Stroud, Sarah. 2014. "Weakness of Will," *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.), http://plato.stanford.edu/archives/spr2014/entries/weakness-will/.

Timmons, M. 1993. "Evil and Imputation in Kant's Ethics," *Jarbuch für Recht und Ethik* 2: 113–41.

Weilenberg, E. 2006. "Saving Character," *Ethical Theory and Moral Practice* 9: 461–91.

Wood, A. 2010. "Kant and the Intelligibility of Evil," in S. Anderson-Gold and P. Muchnik, eds,. *Kant's Anatomy of Evil* (Cambridge, UK: Cambridge University Press): 144–72. .

Zimbardo, P. 2007. *The Lucifer Effect: Understanding How Good People Turn Evil* (New York: Random House).