

## Another Look at Kant and Degrees of Responsibility

PATRICK FRIERSON\*

Whitman College, USA

### Abstract

In “Kant and Degrees of Responsibility,” Joe Saunders claims that “Degrees of responsibility are important for both our moral and legal practices” (p. 1) and argues that “transcendental idealism precludes Kant from vindicating these judgments [about degrees of responsibility]” (p. 11); thus, we have reasons to reject Kant’s transcendental idealism. In this paper, I show how Kant’s transcendental idealism can accommodate and provide a metaphysical account for degrees of responsibility. Whether this “vindicating” such judgments depends upon how much one expects a philosophical account to do; I defend modesty there while admitting a reasonable desire for reflection on how we can and should make such judgments. Finally, I raise the question of just how important judgments of moral responsibility are. Rather than looking to metaphysics to figure out how to vindicate judgments about degrees of responsibility, I suggest we look to the practical purposes such judgments serve.

### Keywords

Kant, Freedom, Responsibility, Transcendental Idealism

Da sprach Adam: “Das Weib, das du mir zugesellt hast, gab mir von dem Baum, und ich aß.”

...Das Weib sprach: “Die Schlange betrog mich, so daß ich aß.”

– Genesis 3:11-13

In “Kant and Degrees of Responsibility,” Joe Saunders claims that “Degrees of responsibility are important for both our moral and legal practices” (p. 1) and argues that

---

\* Institutional Affiliation and E-mail address for contact. Whitman College, [frierspr@whitman.edu](mailto:frierspr@whitman.edu)

“transcendental idealism precludes Kant from vindicating these judgments [about degrees of responsibility]” (p. 11); thus, we have reasons to reject Kant’s transcendental idealism. Saunders offers the example of stealing a chicken:

If you and I both steal a chicken, all other things being equal, we have committed the same wrong and should be held equally responsible. However, if we both steal a chicken, but someone drugged you beforehand, then I am more responsible for this theft than you are. And this responsibility can come in degrees: You could be drugged such that you *totally* lost control of what you were doing, or *mostly* lost control, or *partly* lost control, and so on. (p. 7)

He also considers the cases of children, who are less responsible than adults, and those with mental illness who suffer unspecified conditions “such that their agency is occasionally diminished” (p. 7). According to Saunders, Kant’s transcendental idealism precludes making morally or legally important distinctions in these and similar cases. In this paper, I show how Kant’s transcendental idealism can accommodate and provide a metaphysical account for degrees of responsibility. Whether this counts as “vindicating” such judgments depends upon how much one expects a philosophical account to do, and I defend modesty in that arena, while also admitting a reasonable desire for more reflection on how we can and should make such judgments. Finally, I raise the question of just how important judgments of moral responsibility are. I suggest that rather than looking to metaphysics to figure out how we should vindicate judgments about degrees of responsibility, we should look to the practical purposes that such judgments serve. Importantly, for the purposes of this paper, I am not going to discuss the *legal* practices associated with degrees of responsibility. While Kant can make sense of these practices, the way his philosophy applies to them will be quite different than the way that it applies to *moral* degrees of responsibility.<sup>1</sup>

Before launching into a metaphysical discussion of transcendental idealism, a word about Kant’s moral theory. This theory first and foremost addresses agents deciding what moral demands they are beholden to (along with philosophers seeking to make sense of that agential standpoint). Kant says remarkably little about the assessment of choices that one *has made* or about the choices of others. The categorical imperative is first- or second-person present, an imperative about what to do in order to be a good will. It is not a criterion for determining whether or not a person has or has had a good will. From that perspective, and I’ll return to this briefly in §3, it’s not clear what role degrees of responsibility actually have. When I am trying to decide what to do, I take myself to be responsible; taking myself to be only partly responsible when in the moment of deliberation is a corrupt form of making excuses. In Saunders’s chicken example, if I am

---

<sup>1</sup> Put briefly, Kant’s legal philosophy is based on his Doctrine of Right, where the emphasis is on constructing laws that can secure through coercion a system within which individuals’ actions “can coexist with the freedom [of action] of each in accordance with a universal law” (6:230). Individuals’ motives, ends, and even maxims are not directly relevant to the rightness of their actions under such a system. The question of *legal* responsibility involves the extent to which coercive restrictions on actions under certain circumstances are necessary in order to secure freedom of actions for others.

deciding whether or not take the chicken, and I say to myself something like “Well, I can go ahead and take it, because after all, I’ve been drugged and so am not wholly responsible,” I am guilty of a sort of excuse-making that no moral theory should help to justify.<sup>2</sup>

When it comes to the evaluation of the moral status of individuals, Kant does discuss degrees of imputation, and separately recommends that we “throw the veil of philanthropy over [others’] faults” (6:228, 466; both cited in Saunders, pp. 8-9, 13), but his most sustained philosophical discussion of the ascription of moral status comes in his *Religion*, when he lays out the problem of humans’ radical evil and the fact that we are justified in ascribing evil to all human beings, even the best. In this context, Kant has no qualms about inferring moral status from empirically-given conditions, arguing that “the multitude of woeful examples that the experience of human deeds parades before us” shows universal human evil (6:32-33; for discussion, see Frierson 2013). He also quite clearly is *not* focused on developing an account of degrees of responsibility. In fact, he strongly resists any “lenient justice” (27: 331) or “anything morally intermediate” (6:22), that is, any conceptions of morality that would excuse human beings from recognizing their own evil. Unlike Saunders, who seems quite concerned to ensure that no one is blamed too much, Kant recognizes in human beings a natural tendency to make excuses for our misbehaviors and thereby blame ourselves far too little. His aim in the *Religion*, among other things, is to show how we can honestly come to terms with our own culpability for our own misdeeds while still holding out hope for moral improvement.

Overall, then, Kant just is not particularly concerned with degrees of moral responsibility. Partly, this is because he’s just not that concerned with moral *evaluation* at all, focusing his moral theory instead on articulating one’s duties and giving a philosophical account of the nature and conditions of possibility of taking oneself to be responsible within the practical perspective of deliberation. And when he does focus on moral evaluation, his primary concern is a justification of the claim that human beings are universally evil and a vindication of the possibility of moral hope even in the face of that claim. Fortunately, however, Kant’s vindication of this possibility of moral hope provides resources for thinking through what a Kantian approach to degrees of responsibility would look like. In the following section, I lay out this Kantian approach, before turning in §2 to a discussion of the limits of the approach, and then in §3 to a brief discussion of the actual role of judgments about degrees of responsibility in human life.

### **1. A Metaphysically-Loaded, Transcendental Idealist, Kantian Approach to Degrees of Responsibility**

In this section, I aim to sketch how a Kantian approach to degrees of responsibility could work. For the purpose of this sketch, I draw on a metaphysically thick reading of

---

<sup>2</sup> One might think that there is room for judgments about moral responsibility when it comes to deliberating about how to treat others or how to deal with one’s own misdeeds. I discuss these cases in §3.

transcendental idealism, according to which things in themselves provide the noumenal grounds for phenomenally given objects in the world we experience. More specifically, on this reading human beings have an empirical character in the world of sense that is grounded in a transcendently free, noumenal, intelligible character. Even as a metaphysical reading, this simplifies lots of issues (e.g. whether there are two worlds or two aspects of things, etc.), and in other work of mine (e.g. Frierson 2010), I have argued that metaphysically thin, two-standpoint readings of Kant's transcendental idealism can accommodate many of the claims made on behalf of more metaphysical readings. For the purpose of my discussion above, however, the more metaphysically loaded version allows for a more concise treatment of the problem. It is also, fortuitously, more consistent with the way that Saunders describes Kant's transcendental idealism. As I read it, and I think Saunders could agree with this, the *Critique of Pure Reason* shows, based on the conditions of possibility of empirical cognition, that even the most thorough-going natural necessity could not *fully* explain anything in the world of sense. On this basis, Kant argues that the *ultimate* ground for the empirical characters of things cannot be ascribable to any possible object of experience. This opens room for thinking of transcendently free grounds of events in the world, even while those events in the world proceed in accordance with law-like empirical characters. The *Critique of Practical Reason* then shows, on the basis of the fact of reason – that human beings ought to act in accordance with the moral law – that *we* are such transcendently free grounds. Saunders also rightly draws on and endorses Eric Watkins's reading of Kant, according to which “things in themselves ground appearances, but appearances do not ground things in themselves” (Watkins 2005:328, Saunders p. 5).<sup>3</sup> That is, our law-governed empirical characters are grounded in our transcendently free intelligible characters, but not vice versa. There is no causal influence *from* the world of experience *to* the free grounds of that world.<sup>4</sup>

Now let's throw in some of what Kant adds to this picture in *Religion*. There Kant develops a more detailed model of how humans' intelligible character grounds our actions

---

<sup>3</sup> At times, Saunders is a bit imprecise in the way he puts this idealism, however. Thus he says, for instance, that “every action is either in the world of sense, and thus entirely determined; or outside the world of sense, and accordingly entirely free from such empirical conditions” (p. 6). But for Kant, every action is in the world of sense and thus determined in accordance with natural laws; there's no either-or about this. And every action for which one is morally responsible is entirely free from empirical conditions. As he puts it in the first *Critique*, the “only” question here is “whether it is a correct disjunctive proposition that every effect in the world must arise *either* from nature *or* from freedom, or whether instead both ... might be able to take place simultaneously in the same occurrence” (A536/B564). For Kant, all morally-ascribable actions are *both* determined in accordance with natural laws *and* entirely ascribable to undetermined free agents. This sounds like an impossible combination, which is why Kant had to write two *Critiques* laying out how it can be possible.

I should also note that while I accept Saunders's claim that Kant seeks to reconcile libertarian freedom with strict determinism in the world, some recent Kantians have sought an account of transcendental idealism according to which transcendental freedom precludes strict determinism in the world of experience. Lucy Allais hints at such a view in Allais 2015 and has discussed it with me in conversation. I suspect that she would reject Saunders's claim that Laplace's demon “helps bring out the distinctive nature of Kant's position” (p. 1) and would support a libertarian conception along the lines of Steward 2012 (see Saunders, p. 20, note 43) but would see this as a genuinely Kantian position.

<sup>4</sup> Elsewhere, I've emphasized this point in terms of the “asymmetry in Kant's conception of freedom” (Frierson, 2003: 13ff.)

in the world. Consistent with his long-standing claim that we hold people responsible for the *maxims* of their deeds, and not mere deeds, he claims that intelligible character grounds actions in accordance with maxims. As in the *Critique of Practical Reason*, he explains that ultimately, our maxims are traceable to a fundamental principle, either of self-love or of respect for the moral law. He adds that *all* human beings incorporate *both* self-love *and* respect for the moral law into the most basic maxims of their intelligible characters; the key difference between good and evil, the difference we must ascribe to transcendental freedom in order to make sense of moral responsibility, is the difference between subordinating morality to self-love or vice versa. That is, one is good insofar as one prioritizes the moral law and pursues self-love only within the constraints of morality; one is evil insofar as prioritizes self-love and obeys the moral law only when the cost to oneself is not too high.

Integrating this account of good and evil with Kant's metaphysical picture, we can say that although human intelligible character, as a thing-in-itself, is inscrutable to theoretical reason, practical reason allows us to rationally justify the claim that human intelligible character is characterized by one of these two fundamental maxims (either self-love over morality or morality over self-love). These fundamental maxims cannot be directly experienced, but they show up in the world as empirically-given patterns of choices and actions that are determined in accordance with natural laws. Thus for any given action, we can trace the empirical causes of that action, proceeding "as with any investigation in the series of determining causes for a given natural effect" (A 554/B 582), such as "for a lunar or solar eclipse" (5:99). In addition, we can consider the action morally, tracing it to one or the other fundamental moral maxim.

Two important points here will set the stage for understanding how all this helps with degrees of moral responsibility. First, Kant's argument for transcendental freedom is based on the fact that such freedom is a condition of the possibility of moral obligation or moral responsibility. That means that he has a basis for believing in a transcendently free ground of our empirical character insofar as we hold one another responsible for that empirical character. I generally do not hold myself or others responsible for their hair color. There is thus no reason to ascribe hair color to a transcendently free intelligible character. However, were my friend to dye her hair Home Depot orange, I would hold her at least partly responsible for its color. Importantly, even in that case, there is much that I would *not* hold her responsible for. She's not responsible for the fact that her hair can be dyed, or (at least directly) for the range of colors available to her, or even (importantly) for the particular color that attracted her fancy. But she's responsible for taking her desire (however generated) for a different hair color as a sufficient reason – along with whatever else she took into account – for taking this action. If she colored her hair merely at the behest of Home Depot as a form of advertising, she may even be morally culpable (see 6:423). In short, for Kant, one is responsible only for the *maxim* of one's action, and only for the *form* of that maxim. One is responsible for whether one makes self-love or duty the ultimate basis of maxim endorsement.

And now we come to a second crucial point, the fact of humans' radical evil. This one is more complicated, and could easily become a long tangent, so I'll try to focus only on the basic outline of the story.<sup>5</sup> According to Kant, for most actions looked at in isolation, not only will tracing the empirically determining causes be tricky business, but it is literally impossible to determine which of the two fundamental maxims is the action's ultimate ground. Any action that could follow from a morally good maxim could also follow from a maxim that prioritizes self-love, since one can do good things for bad reasons or because the personal costs of doing them are sufficiently low. However, this inscrutability of underlying moral maxims applies only to *most* actions; there are some actions that *cannot* proceed from morally good maxims – actions that directly violate laws of right, for instance – so in those cases, we can know that the action is grounded in an evil maxim. (For now, I'm bracketing issues about degrees of responsibility, but we'll get there.) Moreover, Kant is a rigorist of a particularly extreme sort when it comes to obedience to the moral law; there is no "middle ground" and "even a single" evil action is sufficient to know that one's underlying maxim is evil (6:20). Thus it seems like each human being must have either a good will that would express itself in a life of moral perfection or an evil will that would express itself in a life that compromises duty for the sake of inclination, at least sometimes.<sup>6</sup> Neither of these intelligible characters would be affected by the empirical world. The will that subordinates the moral law to self-love would be the noumenal ground of an empirical character that takes the form of a temporally-situated willingness to do what's wrong for the sake of satisfying this or that given sensuous (empirical) incentive. These sensuous incentives have no causal influence on the intelligible character; they affect only the way that character shows up in the empirical world. On this account, when we hold someone responsible for their malicious lie, what we are really holding them responsible for is a character by virtue of which they are willing to do what is wrong for the sake of benefit. That character is due to transcendental freedom. The particular way that it manifests itself in this particular case is due to circumstance.<sup>7</sup>

So now we get to what I take to be the central problematic of the *Religion*. Kant shows that human beings universally act in ways that express evil in the fundamental maxims of their wills. This proof of evil in human nature sets the backdrop for Kant's

---

<sup>5</sup> For more, see Frierson, 2003: 95-135 and Frierson, 2013: 72-89.

<sup>6</sup> Neither of these wills would be a "holy will" that is untempted by duty; even a perfect will would be a will that always subordinates non-moral incentives to moral ones, not a will that has no non-moral incentives.

<sup>7</sup> Again, this is a pretty metaphysically loaded way of putting this situation. In two-standpoint terms, what we would say is that from the standpoint of deliberation, one needn't take oneself to be responsible for every specific detail of one's choice situation – say, for the fact that one's inclination is particularly strong, or that it's really easy to get away with in this particular case. What one is responsible for is the way in which one weights those various factors, for one's ultimate values and their relative priorities. That I feel a really strong desire for chicken can be a reason to steal a chicken, insofar I consider really strong desires sufficient bases for violations of right. But that I feel a really strong desire for a chicken cannot, on a Kantian picture, be a sufficient basis for considering really strong desires to be sufficient bases for violations of right. And what I should hold myself responsible for is my choice-architecture, the bases on which I ascribe value, not the particular circumstances in which I apply that choice-architecture. In that sense, I am strictly speaking responsible for only certain aspects of my decisions.

project in *Religion*, which is to articulate a sense of “good will” that could apply to human beings who have good evidence that they are evil. To make a long story short,<sup>8</sup> I take it that Kant replaces a two-fold conception of the options for humans’ intelligible characters with a wider range of possible options. Instead of humans being either morally perfect or morally evil, he allows that a person can be in “revolution” against the evil in her own will. The language of revolution is temporal, which cannot strictly apply to the intelligible will (since that intelligible will cannot be understood with reference to the intuition of time). Kant uses the term “revolution” primarily to indicate that such a will involves evil – since that’s what it is revolting against – but is fundamentally good – since it’s in revolution against evil – and this overall structure, because noumenal, should *not* be understood as partly one way and partly another or as gradually transitioning from one sort of will to another. The idea is that we can have a moral status that is essentially good but that is consistent with at least some evil deeds (e.g., deeds contrary to right) in its empirical expression.

So what would such a will look like? That is, if one’s intelligible character is “in revolution against evil,” what is one’s empirical character like? Kant claims that such an intelligible character would be expressed in the empirical world as a will that does and has done evil, but that is constantly struggling against its own evil tendencies and making progress in that struggle. Crucially, however, this empirical struggle must – to be consistent with Kant’s transcendental idealism – be seen as the *expression* of an underlying noumenal *ground*. One’s will-in-revolution grounds one’s life of struggle and moral improvement, and not vice versa.

There’s a lot more to spell out about this complex picture, and it raises various problems of its own, but I want to focus here on how it can help with degrees of responsibility in Kant. In the context of the problem of human evil, Kant develops an overall model that looks something like this. At the intelligible level, human beings could, in principle, have a will that is one of the following:

- (1) Simply good (prioritizing duty over self-love)
- (2) Simply evil (prioritizing self-love over duty)

OR

- (3) Good by virtue of a revolution (resisting, for the sake of duty, one’s own tendency to prioritize self-love over duty)

These wills look different in the ways that they appear in the world, at least in some cases. While no empirical evidence fully justifies ascribing (2) or (3) to an individual, there is sufficient evidence to know that no actual human beings are simply good (1). Thus a morally good *human* life is a life that expresses a will of the third type, and this manifests itself in a life of constant struggle against evil. But then Kant makes the picture even more complicated, by suggesting that there are actually different sorts of evil, and consequently different sorts of revolution against evil. He describes three sorts (laid out in Saunders’s paper on pp. 18-19): frail, impure, and depraved wills. In his *Metaphysics of Morals*

<sup>8</sup> For longer versions of the story, see Frierson 2003 and Frierson 2013.

(6:407-8) and *Anthropology* (7:251f.), Kant adds yet another sort of “properly evil” will, one that is governed by passion, and he adds a way in which human bodies can be moved that is “merely a lack of virtue” and not properly a form of “will” at all, namely “affects”.<sup>9</sup>

Saunders rightly (in my view) sees these forms of evil as noumenal statuses of the will, ascribable to transcendental freedom. He also rightly claims that “we do not experience transcendental freedom” (p. 19), but he then wrongly infers that “we cannot know whether someone acted out of frailty, impurity, or wickedness” (p. 19). In a strict sense, of course, this is correct. We cannot *know* anything in the moral realm, not even that we are transcendently free; this is a matter for *Glaube*, not *Wissen*. But just as we can have good empirical evidence, in some cases, that human beings are evil, so too we can have good empirical evidence that they are evil of one sort or another. And even when our evidence is shakier, we can assemble enough evidence to make reasonably good assessments. When a person day after day resolves to wake up earlier the next morning and never does so (Kant’s example, see 25:631), it is quite reasonable to see this as frailty, and quite hard to see it as depravity (particularly if one is considering oneself). When over a period of decades someone systematically deceives their customers about the safety of their products, destroys evidence of that deception, and sets up money laundering operations to allow themselves to enjoy the fruits of their spoils without suspicion, it is implausible to ascribe this entirely<sup>10</sup> to anything short of depravity. But the epistemic point is, to some extent, beside the point, for reasons that we’ll get to soon (in §2).

What is central here is that Kant has a metaphysical picture that can make room for different kinds of responsibility for one’s actions. With the exception of affects, all of these forms of evil are aspects of intelligible character for which one is responsible; in that sense, there are not properly “degrees of responsibility” here. But there are quite different *kinds* of responsibility. In every case, what one is responsible for – as noted above – is not the particular deeds one performs, but the structure of one’s volition, the way that one makes use of what is given *to* one in order to decide what to do, and different volitional structures imply different sorts of responsibility.

To see more clearly the different ways that we can hold individuals responsible, let’s compare depraved and frail chicken-thieves. (Stealing a chicken is a good example because, assuming that one knows that one is stealing a chicken, this action violates strict right, so indicates evil.) Someone who is depraved consistently prioritizes self-love over the moral law. Such a person will typically behave quite well, either because of inclinations that tend towards good behavior or because of a social structure that promotes good behavior, or both. For a depraved person to steal a chicken, she would need to have desires that stealing a chicken could satisfy, and she would need to think that she can get away with stealing it without more trouble than it is worth. If we have reasons to think that the chicken-stealing was due to depravity – say, evidence of premeditation – then we hold the thief responsible for consistently and deliberately putting self-love ahead of the moral

---

<sup>9</sup> For extensive discussion of affects and passions, albeit from the standpoint of empirical psychology, see Frierson 2014: 215-58.

<sup>10</sup> Frailty may also be present, but it’s just implausible that the whole system is supported on frailty.



law. The frail person who steals a chicken, by contrast, prioritizes the moral law over self-love in “the maxim of [his] power of choice,” but this maxim “is subjectively ... the weaker” when it comes time to act (6:29). This person commits himself to doing what is right and obeying the law, but in a particular case – say, when starving, or when faced with peer pressure – fails to follow through on his commitment. We do not hold the frail person responsible for deliberately – or at least not consistently and deliberately – putting self-love ahead of the moral law. But we *do* hold the frail person responsible for not holding himself together, for not remaining true to his commitments. And this is precisely what we do *not* hold the depraved person responsible for. Both thieves are moral failures, but in different ways.

I have focused on contrasting two extreme cases, but Kant allows for a much wider range of variation. One can be depraved or frail with respect to different inclinations, or in different sorts of circumstances. Someone who would starve to death rather than steal a chicken might steal it to avoid being laughed at or to get vengeance on an enemy. Someone who is strong-willed under ordinary pressure might be frail when the pressure is sufficiently strong (say, literally starving). Someone who is depraved but cowardly might never steal a chicken but might manipulate others to take the risk for her. And so on. These would be qualitatively different sorts of evil. In some cases, such as those who are frail only under pressures of certain degrees, we might even be able to quantify the differences. Kant himself calls frailty, impurity, and depravity three different “grades” of evil. Thus there is considerable room here for what we might call “degrees” of responsibility. The depraved chicken-thief could be called “more” responsible than the frail one, the frail thief who succumbs to ordinary peer pressure more responsible than the frail one who succumbs to torture, and so on.<sup>11</sup> To that extent, Kant can make sense of degrees of responsibility language, though his moral theory is more conducive to what in my view is a more plausible account, that there are different *kinds* of responsibility. In place of (or, if you like, in addition to) a thin and quantitative scale, we get a rich set of thick responsibility-ascriptions.

Crucially, every case of the ascription of responsibility is an inference, generally imperfect, from given empirical evidence to a description of an empirical character, and from that description to an intrinsically imperfect inference about the noumenal ground that underlies the empirical character. Late one night, I catch Manny – who is usually an excellent neighbor – with my chicken in his hands, jumping the fence from my chicken coop. He immediately gives back the chicken and sobs, confessing that he stole it. He can’t, or won’t, explain why he did so, until gradually I pry from him that he was up late drinking with my consistently troublesome neighbor George, who always has this or that excuse for doing whatever brings him benefit, and who tends to draw others into his trouble-making. Manny immediately takes responsibility for what he did and offers to make whatever amends he can. He says that he’s never done anything like this before – a

---

<sup>11</sup> This implies – pace Saunders – that Kant’s model can quite easily “allow for degrees of responsibility *within* these three different failings” (p. 19).

claim I confirm later – and promises never to do it again. What sort of empirical character does this display? On the basis of only a single incident, it's generally going to be hard to draw conclusions about a person's overall empirical character, but based on the facts as I've described them here, this sounds more like a case of frailty than of depravity. The presence of alcohol suggests that it might even be a particularly mild case of frailty; Manny might have resisted the peer pressure under other circumstances, but he was under the influence, and that probably affected things.

Kant likely would hold Manny responsible for stealing the chicken. Kant does not take frailty to excuse bad behavior; he ascribes frailty to one's intelligible character. Given that we hold ourselves (and others) responsible for failing to follow through on our commitments,<sup>12</sup> such failures must be due to transcendental freedom. In the particular case above, his willingness to so easily go along with a bad influence (both with respect to the drinking and the stealing) is ascribable to him, and his remorse is a partial indication that he recognizes this responsibility. But his evil is an evil of the "not-sufficiently-committed-to-good-to-be-consistent" variety, not the "subordinating-good-to-happiness" variety (nor the impure "not-sufficiently-committed-to-good-to-be-good-without-some-other-reason" variety). Moreover, what Manny is morally<sup>13</sup> responsible *for* is not, strictly speaking, stealing my chicken. He is responsible for having a frail character of a certain type. The fact that he stole my chicken is partly based on that frail character, but also partly based on the character of my neighbor, on the circumstances, the inclinations at play, the effects alcohol has on his inclinations, and so on. Manny isn't responsible at all for most of those things, and for others (e.g. the alcohol) he is only indirectly responsible.

We could considerably enrich this picture by adding the other forms of evil – impurity, and passion – or other details.<sup>14</sup> But only one additional point is necessary before starting to wrap up this section. For Kant, any good human being is good by virtue of being in a struggle against evil. And this notion of struggle introduces another dimension on which we can "rank" different levels of "responsibility." The importance of struggle is already evident in Kant's consideration of the three degrees of evil. One who is depraved is further from good than one who is impure or frail. Someone whose life used to show evidence of unmitigated preference of self over duty (depravity) but who increasingly shows evidence of a commitment to duty that requires the support of additional incentives (impurity) is, on the whole, improving. Such a life, as a whole, reflects a will "in revolution" against evil. More importantly for the present topic, the circumstances of one's actions can affect the reasonableness of thinking that a person is making progress against evil. Take Hu, who grew up in the worst of circumstances, surrounded by vice on all sides, and is drugged by friends and taunted into crime, and then

---

<sup>12</sup> This also implies that if we come to think that there are good reasons for not holding ourselves and others responsible for frailty, we would thereby have good reasons to think that frail behavior is not ascribable to our transcendental freedom. See §2.

<sup>13</sup> Here the distinction between moral and legal responsibility is crucial. Legally, Manny is responsible for stealing my property. That it's a chicken doesn't matter, but it also doesn't matter what sort of motives or frailty he had.

<sup>14</sup> For more on these differences, see Frierson 2014:215-58.

steals a chicken; we can infer almost nothing about her moral status from the theft. For all we know, she is well on her way towards a life of virtue. Since none of us are morally perfect and all are at best struggling against our own evil, evidence of evil cannot be the final word; what we look for is struggle and improvement. That Hu caved under *this much* pressure, given where she started, is no sign that she is not improving. But for Li, who was raised in a family that emphasized virtue, cultivated with good habits and sympathy for others, taught the rewards of justice and doing one's duty, and who is now free of any material needs that a chicken could satisfy, a deliberate plot to steal a chicken simply to cause harm to a neighbor would be a strong – albeit not decisive – sign that he is not struggling against his evil tendencies. We can never know for sure, about any human being, whether they are making progress in virtue or not, but the fact that a *human* good will expresses itself as such progress provides for many ways of distinguishing degrees of moral responsibility.

In principle, there also might be cases in which Manny (or Hu, or Li) would not be responsible at all, even when he or she was the one with the chicken in his or her hands. The most important of these cases, for Kant, are the cases of “affects” and of very young children. Affects, for Kant, are cases where empirically-induced feelings override one's ordinary deliberative capacities to generate behavior that is not controlled by a person's will. Extreme cases of drunkenness might fall into this category, but Kant's more common examples are shock or momentary flashes of rage. In these cases, one is not directly responsible at all. Metaphysically, these would be instances of bodily movement, motivated by sub-rational mental states, that would not be ascribed to one's transcendentally free intelligible character. Just as one does not ascribe one's hair color or heartbeat to one's transcendentally free intelligible character, so too one does not ascribe to one's transcendentally free intelligible character the “acts” of lashing out in a momentary burst of uncontrollable anger or remaining frozen in shock as one's child drowns in a frozen lake.<sup>15</sup> In the cases of very young children (younger than 8-10), Kant does not think that the predisposition to personality – by which we are aware of the binding force of the moral law – has yet been awakened. Such children's actions are not ascribable to their transcendentally free wills.

How far can Kant's overall model go? The metaphysical account here implies that insofar as behavior is part of an empirical character that has as its noumenal ground a transcendentally free actor, that actor can be held responsible for those actions. Anything empirical that does not have such a ground – hair color, inclinations, reflex actions, affects – is not ascribable to a responsible agent. In any given situation, human beings are responsible only for the morally relevant features of their actions, not for the circumstantial features that contribute to the action (including the relevant inclinations). Moreover, they are responsible only when mature (older than eight or ten years old) and when not

---

<sup>15</sup> One might well be held *indirectly* responsible for one's affects. If I have a problem with outbursts of anger and do nothing to rid myself of tendencies to affect, I am responsible for my failure to rid myself of those tendencies.

governed by affects. While one is ultimately either responsible or not, one can be responsible for moral features of one's actions in different ways, ranging from responsibility for bad maxims of choice to responsibility for insufficient self-control. One can be responsible to different degrees, insofar as frailty in the face of torture is "less" frail than frailty in the face of a slight inclination, or willingness to sacrifice duty for the sake of one's life is less depraved than willingness to sacrifice duty for the sake of an extra packet of sweetener in one's tea. Moreover, moral failings for which one is responsible can reflect upon one's ultimate moral status differently depending upon the conditions from which one started and in which one finds oneself. For human beings, a good will is a will struggling against its own evil tendencies, and similar actions can reflect different degrees of that struggle in different people.

## 2. Common Sense, Philosophy, and Degrees of Responsibility

In §1, my goal was to elucidate a Kantian metaphysics for degrees of responsibility. This elucidation shows that Kant's transcendental idealism is not *inconsistent* with a considerable range of judgments about degrees of responsibility. Even if Kant is correct that human beings are transcendently free in their intelligible characters but causally determined in their empirical characters, he can give an account of what it means to say, for instance, that someone who is drugged is "less responsible" than someone who is not, because they are responsible at most<sup>16</sup> for frailty in the face of physiological manipulation. But Saunders often asks for more than merely a metaphysics that is *consistent with* judgments about degrees of responsibility. He repeatedly claims that Kant's metaphysics cannot "vindicate" such judgments (pp. 2, 9, 10n27, 11-12, 13, 18, 19). So I want to spend a little time in this section thinking about just what we should expect from metaphysics in particular and philosophy in general when it comes to judgments about degrees of responsibility.

Let's start with where the account in §1 succeeds at vindicating judgments about degrees of responsibility. At times, Saunders seems to mean by "vindicate" what I claim to have provided above. When he says that "the very notion of diminished agency seems to conflict with Kant's claim that every human action is either entirely determined or entirely free" (7), it sounds like any Kantian metaphysical account according to which one could say that there is a difference between being responsible for deliberately choosing profitable vice over duty (depravity) and being responsible for failing to exercise complete self-control in the face of temptation (frailty) would be sufficient to answer his charge. In that sense, I have "vindicated" Kant in §1. By showing how Kant's metaphysics is consistent with differences in the ways people are responsible for particular behaviors, I have shown how there could be something like degrees of responsibility even given Kant's metaphysics.

---

<sup>16</sup> Here I'm assuming that the drugging does not completely bypass their deliberative capacities. If it does that, then they are simply not responsible, and their actions are no more ascribable to their intelligible character than their hair color or their breathing while asleep.

But Saunders often seems to want philosophy to do more than this. For one thing, he often implies something epistemic, that Kant’s metaphysics should justify not only the claim that there can *be* differences in moral responsibility, but also the claim that we can *know* how responsible someone is (e.g. p. 8). Here Saunders is partly overreading Kant’s inscrutability claims and partly overstating how capable we actually are of making these sorts of judgments. Kant is consistent that we cannot have theoretical knowledge of the moral status of individuals, including whether or not they are transcendently free or to what extent they are responsible. What we have in these cases is practical cognition, rationally justified “belief.” Given the fact of reason, we can have complete rational certainty that human agents, beholden to the moral law, are transcendently free. We can also be sure, given the “multitude of woeful examples that the experience of human deeds parades before us” (6:32), that human beings are evil. About the specific moral status of any individual, including what sorts of evil he is susceptible to, whether or not he is struggling against that evil in ways reflective of a will in revolution, and (therefore) what degrees of responsibility he has for particular deeds, we cannot be certain. We can, however, have good evidence one way or another: “human beings ... can assess themselves and the strength of their maxims ... by the upper hand they gain over the senses in time, ... an ever-continuing striving for the better” (6:48). This evidence will never be perfect. Manny’s sobbing might have been a particularly sophisticated manipulation to avoid punishment, Hu might have taken a deliberate first step in a long term decline into worse and worse depravity, and Li might be acting out in a first step towards overcoming deep-seated but previously hidden wicked maxims. But those epistemic limitations when it comes to the evaluation of others are just real facts about our human condition. Not all fourteen year olds are less responsible than adults (to use Saunders’s example, p. 7), and it’s not always easy to tell when they are. Ascriptions of degrees of responsibility are judgment calls.

Saunders also wants philosophy to do more in the way of vindication in another respect; he often seems to want philosophy to *justify* our judgments of degrees of responsibility. Thus when he criticizes my claim that “the method for determining what the appropriate markers are starts with common-sense moral judgements about responsibility” (Frierson 2012:186, quoted on p. 11), he says, “However, this is not a solution – it is the very problem at hand. The problem is that transcendental idealism precludes Kant from vindicating these judgements in the first place” (p. 11). When criticizing Korsgaard’s view, he asks, “From the practical standpoint, how can we distinguish between a person and a toddler? It is not enough to say, ‘we just do’. This sidesteps the important issue of how Kant can vindicate these practices.” (p. 12). If “vindicate” just means “give a metaphysics consistent with,” then Kant can vindicate these practices; toddlers’ empirical characters are not grounded in transcendently free intelligible characters, while those of adults are. It seems that what Saunders really wants here is a transcendental idealism that will show that X or Y markers of moral responsibility

are the correct ones, and that X or Y practices of holding people responsible in degrees are justified.

I understand why Saunders wants this. It seems, intuitively, like the issue of moral responsibility should have a metaphysical ground. It seems like in order to know whether or not someone is responsible, we should need first to figure out some metaphysical (and probably also psychological) facts about the relationship between that person and the action at issue, and then make a philosophically grounded inference from those facts to ascription of responsibility. It seems like metaphysics should come first, and practical philosophy should follow as a consequence of what we discover. But that's not Kant's view. For Kant, the "primacy of the practical" implies precisely that in a whole sphere of metaphysics – the traditional problems of freedom, immortality, and God – practical reasoning comes first, and we draw metaphysical conclusions by considering the conditions of possibility of practical claims. At least in the *Critique of Practical Reason*, Kant argues *from* the fact that human beings are obligated by categorical imperatives *to* the fact that we are free. He argues *from* the fact that we ought to promote the highest good *to* the fact that such promotion is possible (and therefore that we are immortal and there is a God). His metaphysics does not prove his practical philosophy. Proving that we are immortal does not prove that we have an obligation to become perfectly virtuous. Rather, the claim that we have an obligation to become perfectly virtuous proves the metaphysical claim. But the metaphysics "vindicates" the practical philosophy in a weaker sense. It shows that there *are* possible conditions of possibility of the practical claims we are bound to make. That is, it shows that those (necessary) moral claims are not impossible; there is a metaphysics that can support – in the sense of being consistent with – them.

As I read him, Kant starts both his theoretical philosophy (in the first *Critique*) and his practical philosophy (in the *Groundwork* but especially in the second *Critique*) from common sense claims (cf. Ameriks, 2000). In the first *Critique*, he takes for granted that we have experiences of ordinary empirical objects and that we legitimately make a priori synthetic judgments in mathematics and natural science, and he develops a metaphysics that shows how it's possible for us to do that. In the second *Critique*, he takes for granted that we have moral obligations, and he develops a metaphysics that shows how this is possible. Combined, the two *Critiques* show how the common sense claims about the empirical world are consistent with common sense claims about morality, and in particular how the causal determinism that is a condition of possibility of the practice of natural sciences and even ordinary temporal experiences of objects is consistent with the transcendental freedom that is a condition of possibility of morality. But all of this "vindicates" common sense not by providing a foundational proof that various common sense practices are correct but by laying out a metaphysical picture according to which they are not necessarily impossible.

What I've defended in §1 is a vindication of this sort. I've shown that Kant's metaphysics allows for distinguishing different sorts of intelligible characters based on the sorts of evil to which they are committed or against which they are struggling, and I've shown how this can map on to our ordinary practices of judging different degrees of

responsibility. But these claims about different sorts of radical evil, and particularly the claim that they map onto different degrees of responsibility, depend upon rather than independently prove the relevant common sense judgments about such responsibility. This dependence is clearest in the cases where we do not ascribe responsibility at all. Toddlers and those in momentary fits of rage are not treated as morally responsible for their actions, *and so* we do not ascribe those actions to their intelligible character. We do not first find a metaphysical basis for not ascribing the actions to their intelligible character *and then* infer that they are not responsible. If that's what Saunders wants in the way of vindication, he's not going to get it from Kant's philosophy.

The reason he's not going to get that from Kant is one to which I drew attention in the article of mine with which Saunders engages most. For Kant, philosophers in general and metaphysicians in particular are very good at certain highly refined sorts of reflection, and they can have important roles to play in coming to better understand how various commitments fit together into a coherent whole. In morals, philosophers can even help combat tendencies to self-deception that are pervasive in ordinary life and that drive people to mitigate the demands of the moral law on themselves. But, as Jeanine Grenberg aptly emphasizes in the title of her book, Kant's philosophy as a whole, and his moral philosophy in particular, is a "*Defense of Common Moral Experience*" (Grenberg, 2013). Kant starts with common sense and shows how it could possibly be true; he defends it against any "reasonable doubt" about it. He's not trying to prove it afresh.

### 3. Why do we need degrees of responsibility?

I want to end here in a similar place to where I ended in the 2012 article to which Saunders responds, but say a bit more about how we might move forward, in the context of degrees of responsibility. In that 2012 article, I pointed out that "Kant's philosophy needs to be supplemented with an account of how one can arbitrate between competing common sense views of moral responsibility" (Frierson, 2012: 482). Since Kant doesn't do the common sense work for us, and he doesn't replace common sense with metaphysics, we need to think about how we might refine our understanding of degrees of responsibility given Kant's overall metaphysical vindication (in *my* sense of vindication) of them.

It's important to start here with three fairly obvious but very important and deeply Kantian points about degrees of responsibility. First, as Saunders recognizes in his own article, many ascriptions of degrees of responsibility are not difficult in ordinary life. While there may be "hard borderline cases," there are plenty of "clear case[s] of someone having more or less ... responsibility" (p. 7), such as the differences between adults and children, those under the influence of mind or mood altering drugs and those not under such influences, and those with certain sorts of mental illnesses. Insofar as there are straightforward common sense views about degrees of moral responsibility, we can legitimately use Kant's metaphysical views about the nature of human freedom in order to make sense of – and thereby "vindicate" in a weak sense – those common sense views.

Second, while there are some tough cases – as Saunders notes, “judges and juries navigate a difficult terrain” – the difficulties in these cases are typically not metaphysical or even ethical but much narrower and less philosophically interesting. We need to know not whether, for instance, being drugged is relevant to responsibility, but about whether and to what extent and with what effects a person really was drugged. We need to know not whether it matters that she used that insulting term without intending to offend but whether (and how, and to what degree) she understood that it was insulting. These can be genuinely hard issues to figure out in particular cases, but nothing about transcendental idealism will make them harder (or easier). Finally, a point Saunders wholly ignores, it’s important to acknowledge that many judgments about degrees of responsibility are used – illegitimately – in self-exculpatory ways: “we like to flatter ourselves by falsely attributing to ourselves a nobler motive” (4:407). Even at a common sense level, we can diagnose that self-exculpation is a moral problem, and we can be suspicious of demanding too much in the way of a philosophical justification of excuses for one’s own bad behavior.

Still, however, we may need some ways of thinking about when judgments about degrees of responsibility are legitimate, in cases where the relevant empirical facts are clear but the consequences for degrees of responsibility are not. And Kant’s metaphysics cuts off at least one natural way of drawing this line. It can’t be the case that our degree of responsibility corresponds to the extent to which our action is determined by causal laws of nature. Since every action is determined in accordance with causal laws of nature, such a rule for drawing the line would absolve everyone of responsibility for everything they do.<sup>17</sup> And while we can legitimately say that we are responsible for all and only those actions that are ascribable to our transcendently free intelligible character, and responsible only for the contributions that our free character makes to those actions, this metaphysical account doesn’t help us pick out what are responsible for, since we decide what is ascribable to our intelligible character only by inferring the metaphysics *from* the ascription of responsibility. So we need a different sort of reflection than the sort that Kant gives us in his transcendental idealism. And this reflection can’t be metaphysical; Kant’s *Critique of Pure Reason* rules out the possibility of having theoretical insight into human freedom. In this arena, we must go *from* practical reasoning *to* metaphysics, not vice versa. So we need some better sort of practical reasoning.

Here I’d like to make a conjecture that’s not particularly well grounded in Kant, though it has some resonance with his claim that we should “throw the veil of philanthropy over [others’] faults” (6:466)<sup>18</sup> and had been prompted by my own reading of what Kant is

---

<sup>17</sup> Alternatively, since nothing in nature is *sufficiently* determined in that one can always ask for further explanations of the further grounds for any causal laws governing any particular instance, we might say that everyone is wholly responsible for everything. (“Granted, you stole the chicken because the chip in your brain overrode your decision and moved your body so that it picked up the chicken, but why was the chip in your brain able to do that?”)

<sup>18</sup> Incidentally, it’s worth noting that the context of this remark involves yet another warning by Kant against self-exculpatory behavior. What he specifically condemns is taking “malicious pleasure in exposing the faults of others *so that one will be thought as good as ... others*” (6:466, emphasis added).



up to in his *Religion*.<sup>19</sup> I'd like to suggest a Kantian constructivist approach to degrees of responsibility.<sup>20</sup> Instead of taking judgments about degrees of responsibility to be grounded in some metaphysical truths about the world, let's take them to be grounded in practices of holding responsible, practices that we construct in order to meet needs that arise within practical life together. In this context, rather than simply taking for granted that they are "important" (Saunders, p. 1, 8) let's ask "What do we need judgments about degrees of responsibility *for*?" And again, recalling that we often use such judgments illegitimately, in order to excuse the bad behavior of ourselves or of those we particularly love, the real question is "What *morally legitimate* reasons do we have to make judgments about degrees of moral responsibility?"<sup>21</sup> So here's a partial list of some *prima facie* legitimate reasons we might make judgments about degrees of responsibility:

1. To determine whether and how much punishment is appropriate
2. To determine how much praise or blame is due to a person
3. To determine how much remorse or regret one should feel for a given action
4. To decide whether forgiveness is appropriate in a given case.<sup>22</sup>
5. To give oneself or another moral hope, a sense that X or Y misdeed does not preclude virtue altogether
6. To promote effective moral cultivation in oneself or others

For determining the appropriateness of punishment, blame, forgiveness, remorse, and hope, and for discerning the best ways of cultivating moral improvement, it can make a difference just how responsible one holds a person to be. A young child will not deserve (as much) punishment and moral blame, and should not feel the same degree (or perhaps even kind) of remorse and regret. Strictly speaking, it doesn't make sense to "forgive" a child for a wrong for which they were not responsible. And with very young children, promoting moral cultivation is not primarily a matter of getting them to commit to the right moral principles as it is a matter of cultivating their abilities of self-governance in general. By contrast, fully functional adults making decisions in the context of careful deliberation and with ample self-control are rightly held fully responsible, subject to punishment, blame, remorse, and so on.

<sup>19</sup> The details of how this relates to the *Religion* are a topic for another paper.

<sup>20</sup> Or, perhaps, a *post*-Kantian constructivist approach. In some ways, the basic outlines of this approach fit better with Hegel than Kant. Cf. Pippin, 2008. In other ways, it fits better with pragmatists, from James to Rorty and Brandom.

<sup>21</sup> Here, in particular, it's important to recall that Kant's ethics is primarily about deliberation about what to do, and it's worth noting that consideration of degrees of responsibility is almost always illegitimate in such contexts; if I am asking, "How responsible am I for what I am about to do?", then I'm in moral trouble. This anticipatory self-exculpation is a form of radical evil, a way in which we corrupt our own principles of volition. And it is a classic example of what Kant warns against in *Perpetual Peace*, a self-fulfilling negative prophecy, where precisely because I take myself not to be able to do anything but what's wicked, I don't do anything but what's wicked.

<sup>22</sup> Here my thought is that forgiveness is *not* appropriate in cases where a person is not responsible for what they did, and only appropriate to some degree when a person is only responsible to some degree. We can exculpate or set aside the bad deeds of someone who is not responsible, but we cannot *forgive* someone that we don't hold responsible. For discussion of related issues, I'm grateful to Lucy Allais.

A full discussion of each of these different reasons for making judgments about degrees of responsibility would be far beyond the scope of this paper. Here I just want to make a few short remarks about the practical reasons for such judgments and the metaphysical requirements for them. Of the reasons listed above, by far the most prominent is the use of degrees of responsibility in legal contexts. For Kant, however, legal punishment does not depend upon the moral condition of the criminal: “The concept of right ... has to do ... only with the external ... practical relation of one person to another, insofar as their actions, as deeds, can have direct or indirect influence on each other” (6:230). The only relevant question when it comes to punishment is whether a given punishment is “a hindrance to a hindrance to freedom,” (6:231), that is, whether the punishment functions to counteract an action that hinders legally permissible actions. There might be various reasons for taking into account exculpatory circumstances in determining legal sanctions for actions that are inconsistent with the freedom of everyone in accordance with a universal law” (6:232), but the relevant “degrees of responsibility” will not require any deep metaphysical underpinnings.

With respect to the other reasons for making judgments about degrees of responsibility, there is also no reason to think that such judgments require deep metaphysical commitments. Kant makes clear in his *Groundwork* that “praise and encouragement” do not depend upon “true moral worth” (4:398). Someone with good inclinations – for which nature, not one’s transcendently free intelligible character, is responsible – nonetheless gets praise. Whatever the reasons for doing this, they do not depend upon a metaphysical foundation. So why praise or blame in the absence of knowledge of a person’s moral worth? Praising others for virtues that they may or may not have can also serve a pedagogical role and a role in socially reinforcing virtue, both through clarification and endorsement of what sorts of actions are genuinely in accordance with the moral law and through enlisting supporting motives for virtuous action, such as the love of honor that Kant calls a “simulacrum of virtue” (2:218). Praise of others can also help combat misanthropy and thereby promote genuine respect for others. Blame, too, can serve some of these pedagogical roles by showing what sorts of actions are blameworthy and thus to be avoided and by enlisting the love of honor as a support for moral incentives, but excessive moral blame of others risks fostering misanthropy, which can promote disdain for others, self-righteousness, and/or moral despair, all of which inhibit one’s ability to treat others with respect. Thus Kant recommends that “the veil of philanthropy” be thrown over others (6: 466), and ascribing to others a diminished degree of responsibility for their faults is one way to throw such a veil.

Strikingly, almost none of these valuable functions of praise and blame actually depend upon metaphysically-grounded or even upon accurate ascriptions of degrees of responsibility. When we praise someone for good behavior, what matters is how much we want to endorse and encourage that behavior, not how responsible they are. And in the one case where degrees of responsibility do seem most helpful – the softening of our blame of others – Kant’s own recommendation is that we avoid the “offensive inquisitiveness” that comes with “spying on the morals of others” (6:464). That is, rather than taking it upon

ourselves to judge that Manny or Hu is not really morally blameworthy because not wholly responsible, we should simply soften our anger towards them out of a recognition that such anger corrupts the respect we owe to them as people, and then refrain from making judgments about their fundamental moral convictions. We should, in other words, judge people *less*, rather than trying to make more and more fine-grained judgments of just how responsible they are.<sup>23</sup>

With respect to ourselves, however, “the duty ... is ... to cultivate one’s conscience, to sharpen one’s attentiveness to the voice of the inner judge and to use every means to obtain a hearing for it” (6:402). In one’s own case, we should *not* refrain from moral scrutiny. Nonetheless, even here, there is little reason to insist upon degrees of responsibility. If I have violated the moral law, I should feel regret and remorse, even if I was not wholly responsible. Marcia Baron explains that “remorse...involves a judgment on the agent’s part that she acted wrongly and should and could have acted differently, ... [and] the judgment is action-guiding; ... It has motivational force” (Baron 1988: 259). One who feels remorse should reform his ways, and often should deal with relevant moral remainders. If I betrayed your secret during a bout of frivolous gossiping, I owe you at least an apology, and I may need to take concrete steps to help mitigate the harm to you done by the betrayal. If through careless neglect or deliberate malice I ruined a book that you lent me, I owe you a new book. In both cases, I should feel remorse or regret for my bad actions and commit myself to not repeat them in the future. What difference does it make how “responsible” I am for those bad actions? On the one hand, suppose, in the most extreme case, that I am not responsible at all, that – to use Frankfurt’s wild scenario – a chip was implanted in my brain that made me do it. I still ought not maliciously gossip or destroy books in the future (in cases where I do have self-control), and I ought to take steps to increase my self-control. In those senses, I should feel regret. Moreover, since, after all, you were still hurt, I should take steps to remedy your hurt. On the other hand, supposing the opposite extreme, where I deliberately planned the revelation of the secret or destruction of the book, and I did so knowing that it was wrong, and just for the sheer sadistic pleasure of harming you, a pleasure that I used not to feel but a susceptibility for which I have deliberately cultivated because I concluded that it would be a particularly sweet pleasure and contribute well to my overall happiness. In this case, I should feel

---

<sup>23</sup> In private comments on an early draft of this paper, Saunders asks the seemingly reasonable question, “Why not both?” In my view, however, the options are by and large mutually exclusive in actual practice. Generating fine-grained judgments of moral responsibility requires paying close attention to just what a person is and is not responsible for in a given case (usually a case of a given misdeed). This requires spending more time judging the actions of others. Correlatively, given the reality of human evil, those who spend a lot of time judging others typically end up either isolated and misanthropic or end up needing to absolve others through ascriptions of diminished responsibility. Of course, there will likely be cases where there is good evidence that others are morally corrupt and we cannot help judging them in some way (as, perhaps, in the case of Manny?). In those cases, beyond keeping our judgments as much as possible to ourselves (6:464), Kant exhorts us to adopt a posture of *forgiveness* rather than *exculpation*, for reasons I’ll briefly mention below. When there is strong evidence of diminished responsibility, we can of course attend to such evidence (and §1 shows how empirical features could evidence such diminished responsibility), but we do better to refrain from judgment where possible.

remorse, I ought to commit myself not to engage in this behavior in the future, and I ought to provide you some recompense for the harm I've caused.<sup>24</sup>

I'll deal with the last three reasons to use degrees of responsibility more quickly, although they really deserve the most extensive treatment. Given that human beings are evil, and in that evil cause harm to one another, there is a serious danger of moral despair. Our relationships with others can seem irretrievably broken due to their evil actions towards us, and real hope in the moral goodness of ourselves and those we love can erode in the face of the "multitude of woeful examples" that experience "parades before us" (6:33). Judgments about degrees of responsibility are, among other things, ways of coping with these broken relationships and moral despair. By thinking that a person was not wholly responsible for an action, we can cut them some slack and hope for better times in the future. As I explained in §1, there is room for this sort of exculpation of moral wrong in Kant's metaphysics. At the most extreme case, I can see my lover's betrayal as something for which he is *wholly* non-responsible, a mere behavior motivated by affect. But I can also see it as mere frailty, perhaps in the fact of extraordinary circumstances, rather than depravity or impurity. And so on.

As appealing – and common – as it is to appeal to diminished responsibility in these cases, however, there are important reasons Kant does *not* turn to degrees of responsibility as his primary way of dealing with human evil. Kant thinks that he has an alternative account, a different way of restoring relationships and justifying moral hope. Though he does not wholly work it out, Kant turns to "grace" and "forgiveness," rather than lenient justice or degrees of responsibility. We can and should accept that ourselves and others really are evil, but we also can and should accept that forgiveness is possible, that despite evil we can move forward. Grace and forgiveness, whatever their problems, are more respectful of humans' dignity and freedom than exculpation. And when we are unsure, in a given case, how responsible a person is, forgiveness provides a way of moving forward from misdeeds that admits that whenever there is *any* responsibility, forgiveness can be justified. A community within which people more readily forgive is one within which judgments about degrees of responsibility have much lower stakes.<sup>25</sup>

---

<sup>24</sup> Leaving aside issues of legal right, degrees of responsibility don't even change the *kinds* of obligations that I have in the light of harm to others. However responsible I was, I have a perfect obligation not to repeat the bad actions and imperfect obligations to remedy the harms I have caused and to cultivate in myself the tendencies (self-control, sympathy, etc.) that will make it more likely in the future that I'll behave well. It may be that my indirect duties should carry more weight if I was more responsible for what I did, but I suspect that how responsibility affects subsequent responsibilities will vary considerably on a case-by-case basis. Because Kant talks so little about moral (as opposed to legal) recompense, it's hard to know precisely how he would deal with different sorts of obligations, but the basic structure of his moral theory suggests that the moral responsibility of the actor will be at best only an indirect factor. What matters most fundamentally is the humanity of the persons principally affected, and the casuistical questions here are issues of how best to respect others' (and one's own) humanity in the context of violations of various kinds (where, for instance, the harm of missing a chicken is different than the harm of being undeservedly treated as an enemy by a neighbor).

<sup>25</sup> Note, again, that I am talking here about moral community, not legal community. There may be good reasons that legal punishment is required even when moral forgiveness is proffered.

Finally, there is the issue of moral cultivation. Here, too, degrees of responsibility play much less of a role than they initially seem. There are quite complex issues about how moral cultivation can work at all in the context of transcendental idealism; I've discussed these in detail in Frierson, 2003. But insofar as those issues can be resolved, what matters is what empirical influences provoke actions that conflict with the moral law. If one has a frail will, one needs to cultivate strength of will; if one is deprived, one needs to restructure the principles on the basis of which one chooses. If one is primarily tempted by animalistic inclinations, one needs to work specifically on resisting those inclinations; if one is just generally inconstant in one's behavior, one needs to work more generally on constancy. And so on. Some of these cases may involve different degrees of responsibility than others. But the degrees of responsibility don't affect how one would actually work for self-cultivation. If one is afflicted by affects, and so not responsible at all, he still has a responsibility to rid himself of affects; if afflicted by a passion, and so deeply responsible, she still has a responsibility to rid herself of passion. How one works against passion is quite different than how one works against affect, so it matters what the causes of our bad behavior are for how we work against them, but the fact that we have more responsibility in one case than another doesn't affect our practical endeavors.

Overall, then, there are many uses of judgments of degrees of responsibility – particularly for self-exculpation – that are unnecessary or even dangerous. In other cases, judgments about degrees of responsibility are second-best attempts to do what can be done better. We can soften our judgments of others better by *refraining* from judgment, and we can deal with others' misdeeds better through forgiveness, than through ascribing to them diminished responsibility. Finally, there are judgments of degrees of responsibility that are perfectly appropriate but can be justified for pragmatic reasons that don't require any profound metaphysical backing. Even where degrees of responsibility are important, where some metaphysical backing is called for, Kant's account of human beings is sufficient. Human beings are radically evil but potentially in revolution against their evil, where such a revolution would constitute a human good will and show up empirically as a struggle against one's own evil tendencies. We give people more praise for actions that more clearly demonstrate this struggle against evil, and we partially excuse behavior that, in the context of the person's life as a whole, provides evidence more of universal human propensities to evil than of a failure to struggle against those propensities. The proper judgments here will more often be qualitative than quantitative, primarily judgments about kinds of responsibility and secondarily (if at all) about "degrees." These judgments will also be imperfect, both because empirical evidence of what happened is imperfect and because we can only imperfectly infer fundamental (noumenal) moral character from the empirical expressions of that character. Kant's transcendental idealism thus provides a metaphysical picture consistent with different kinds and degrees of responsibility, while his epistemology and practical philosophy set realistic limits on what we can judge about others and on how much (and how) we should judge them.

## References

- Allais, L. (2015) *Manifest Reality: Kant's Idealism and His Realism*. Oxford University Press.
- Ameriks, K. (2000) *Kant and the Fate of Autonomy*. Cambridge University Press.
- Frierson, P. (2003) *Freedom and Anthropology in Kant's Moral Philosophy*. Cambridge University Press.
- Frierson, P. (2010) "Two Standpoints and the Problem of Moral Anthropology," in *Kant's Moral Metaphysics* (ed. James Krueger and Benjamin Lipscomb), Berlin: Walter deGruyter Press, pp. 83-110.
- Frierson, P. (2013) *Kant's Questions: What is the Human Being?* Routledge.
- Frierson, P. (2014) *Kant's Empirical Psychology*. Cambridge University Press.
- Grenberg, J. (2013) *Kant's Defense of Common Moral Experience: A phenomenological account*. Cambridge University Press.
- Pippin, R. (2008) *Hegel's Practical Philosophy: Rational Agency as Ethical Life*. Cambridge University Press.
- Saunders, J. (2018) "Kant and Degrees of Responsibility," in *Journal of Applied Philosophy*.

