

Kant on Radical Evil and the Origin of Moral Responsibility

IRENE MCMULLIN
University of Arkansas
Email: mcmullin@uark.edu

Abstract

The notion of radical evil plays a more important role in Kant's moral theory than is typically recognized. In *Religion Within the Limits of Mere Reason*, radical evil is both an innate propensity and a morally imputable act – a paradoxical status that has prompted commentators to reject it as inconsistent with the rest of Kant's moral theory. In contrast, I argue that the notion of radical evil accounts for the beginning of moral responsibility in Kant's theory, since the act of *attributing* radical evil to one's freedom is an inauguration into the autonomous stance.

Keywords: autonomy, obligation, incentive, innate propensity, intelligible act, maxim, moral law, radical evil, respect, responsibility

Introduction

Kant's notion of 'radical evil' is notoriously problematic. Ostensibly, Kant is claiming that human beings are simply born with an innate propensity to prefer acting on inclinations of self-love to those of the moral law. Despite characterizing radical evil as an innate propensity, however, Kant also claims that it is *chosen* – a position he must endorse if he is to hold the agent responsible for her failure to act on the moral law. In response to such a paradoxical position, many commentators have rejected Kant's notion of radical evil as an unhelpful addition to his moral theory, suggesting that this notion is a holdover of the religiosity of Kant's life and times or else reducible to an anthropological assertion.¹ Most commentators agree that the notion of an innate radical evil seems at odds with Kant's moral theory insofar as he places morality squarely in the noumenal realm whereas claims about 'human nature' belong to the phenomenal – a position that makes it

difficult to see how humanity is by *nature* disposed toward an evil for which it is *morally* responsible.

My purpose is to show why the notion of radical evil plays a more important role in Kant's moral theory than is typically recognized. I argue that the notion of radical evil accounts for how the agent first comes to take responsibility for the fact that she is claimed by the moral law and yet is inclined to prioritize other incentives. By making herself answerable for this innate distance between incentive and law – a responsibility-taking that first occurs when she attributes radical evil to her freedom – she is transformed into a genuinely moral agent. Thus 'radical evil' must be understood as a *prescriptive*, not merely a *descriptive* self-designation in the sense that one *ought* to interpret oneself according to that category if one is to become responsible for one's moral condition. Only through such an act of self-attribution can we make ourselves responsible for our moral failures – despite the fact that we did not *choose* to be the imperfect creatures that we are.

Are We Evil?

In *Religion Within the Boundaries of Mere Reason*, Kant attempts to account for the fact that one's will does not correspond perfectly to the moral law but is rather characterized by a type of inner conflict whereby the dimensions of obligation and response diverge within the will itself. In the human will (contrary to the holy will) nothing guarantees the coincidence of the moral law, the *objectively obligating* moment of the will, with incentive, the *subjectively responsive* moment of the will, which is responsible for harmonizing itself with, or contradicting, the moral claims made upon it. This disjunction, Kant claims, is due to the fact that the subjective determining ground of the will can be either the moral law *or* inclinations of self-love. Though the moral law always obligates unconditionally, the will is nevertheless determined to action by various incentives:

The law rather imposes itself on him irresistibly, because of his moral predisposition; and if no other incentive were at work against it, he would also incorporate it into his supreme maxim as sufficient determination of his power of choice, i.e. he would be morally good. He is, however, also dependent on the incentives of his sensuous nature because of his equally innocent natural predisposition, and he incorporates them too into his maxim (according to the subjective principle of self-love). (Kant 1996: 82–3; *R* 6: 36)

Human agents are torn by conflicting incentives – moral and sensuous – and though we are not responsible for the *presence* of the ‘innocent’ sensuous incentives that compete with the moral law, we *are* responsible for allowing them to trump it.

Radical evil refers to humanity’s propensity (*Hang*) to allow the sensuous incentives to take pride of place. More exactly, it refers to humanity’s tendency to make conformity to the moral law conditional on agreement with incentives of self-love. To be ‘radically evil’ is not simply to fail to be moral on this or that occasion – it is for these specific instances of immorality to indicate a general tendency to make moral incentives dependent upon their agreement with sensuous ones, rather than placing incentives of self-love second to those of morality. Despite the fact that Kant acknowledges the possibility of an analogously innate ‘radical’ good, he nevertheless pronounces judgement on all humanity by claiming that it is *universally* characterized by such an innate disposition to radical evil – the disposition to make moral choices conditional on incentives of self-love (Kant 1996: 79–83; R 6: 32–6).

But characterizing this propensity as *universally* innate seems to imply that Kant slides into a Pietist misanthropy inconsistent with the moral optimism of his other texts.² This conclusion appears to be justified in so far as Kant gives us little reason to agree with his assertions about the universality of this innate propensity. He simply claims that the ‘multitude of woeful examples’ (Kant 1996: 80; R 6: 32–3) allows us to conclude that human nature is indeed universally tainted by this propensity to allow self-love to trump the moral law.

Despite such casual empirical claims about its universality, Kant also indicates that a formal (i.e. *a priori*) proof for the necessity of this condition is possible, though he himself does not provide it. Several scholars have attempted to fill in the gap underlying Kant’s ‘breezy assertion’ that such a proof is possible (though not forthcoming).³ The problem with such accounts, however, is their tendency to mischaracterize the role that the notion of radical evil plays in Kant’s project by understanding it as a *descriptive*, rather than a *prescriptive* notion. Contrary to such approaches, I will argue that ‘radical evil’ is primarily a backward-looking self-designation that allows the agent to make sense of her moral experience. It is not a *description* of practical reason’s structure, but rather a self-interpretation of that structure that is necessary for the proper functioning of practical reason.

To see why this is the case, we must examine the seemingly contradictory requirement that radical evil be understood both as an innate propensity *and* as originating in an act of freedom. Though the tendency to disorder the incentives is taken to be part of human nature, it is also characterized as somehow arising out of free choice. Kant must reconcile the seeming inconsistency of these two claims by grounding humanity's universal innate propensity to evil in free choice. If he cannot accomplish this, the disordering of the incentives could only be understood as 'a natural impulse' (Kant 1996: 71; R 6: 21) and thus not something for which we are responsible. Because Kant wants us to be culpable for this failure to conform to our moral obligations – and we are not culpable for *natural* impulses – the prioritization of self-love over the moral law cannot be attributed to a natural cause. If we are to take responsibility for it – as he believes we should – the propensity to invert the appropriate hierarchy *must* be due to an act of freedom.

The inconsistency in Kant's position appears particularly evident in so far as acts of freedom *cannot* be traced to some ultimate subjective ground – as such an 'innate propensity' would be. The ground of freedom is 'inscrutable' (Kant 1996: 71; R 6: 21) because all attempts to trace its origin 'further back' result either in (illicit) natural explanations or an infinite regress of maxims. Acts of freedom – qua *free* – are a type of pure or unconditioned original beginning that can only be traced back 'so far' before we are in danger of mischaracterizing them as events in the causal order (Kant 1996: 80, 85–6; R 6: 32, 39–40). An act of freedom is not a mere event that happens to originate with an agent, but is, rather, an act governed by a maxim that incorporates incentives in a principled way. Acts of freedom *must* be governed in this way if they are to avoid sliding into pure arbitrariness – an arbitrariness that undermines the possibility of genuine choice.⁴ Because acts of freedom are not causally determined but rather normatively governed events, accounting for their origin in causal/natural terms is incoherent. Thus if we are imputable for our failure to live up to the moral law, this failure must arise from an act of freedom, but 'apart from a maxim no *determining ground* of the free power of choice ought to, or can, be adduced' (Kant 1996: 71; R 6: 21, footnote).

Kant's claim that radical evil is an innate propensity seems to violate exactly these conditions, however, by reducing a free act to a kind of natural fact. In what sense is the choice of self-love over morality both 'free' *and* an unchosen tendency that '*precedes every deed*, and hence is itself not yet a *deed*?' (Kant 1996: 79; R 6: 31)? How can there be an

origin that we did not explicitly choose but for which we are nevertheless responsible?

Intelligible vs Empirical Acts⁵

The answer lies in Kant's distinction between intelligible and empirical acts. According to Kant, what characterizes intelligible acts is that they are 'cognizable through reason alone apart from any temporal condition' (1996: 79; R 6: 31). When considering something to be an intelligible act, the focus is on the *existence* of the *effect*, not on the empirical event that is taken to have brought this effect into being (Kant 1996: 85; R 6: 39). Empirical acts, on the other hand, refer to all those exercises of freedom whereby specific material actions are performed at identifiable points in time. Thus intelligible acts can be understood as *acts* in the sense that they arise out of and are attributable to one's freedom, but unlike empirical acts, they cannot be understood as having so arisen because of a particular *event of choice* locatable at a specific point in time. This distinction is essential for understanding how Kant can claim that radical evil is an 'act' – i.e. attributable to one's freedom – and yet deny that it involved an explicit event of choice. The origin of the propensity to evil is an 'act' in the sense that it is an exercise of freedom, the choice of a supreme maxim according to which one either harmonizes one's will with the moral law or contradicts it. What is initially interpreted as an innate disposition – since it is a 'given' that we do not choose at a particular point in our lives – is now interpreted as an intelligible act. And because it is an *intelligible* act, only the *consequence* of being 'in' a condition of moral failure – and not the *event* whereby the 'choice' of misordered incentives was supposedly made – can be made accessible to reason. Thus the fact that we sometimes choose self-love over reason is made comprehensible to reason by interpreting this fact as having arisen from a general choice to make moral acts dependent on their agreement with self-love – a choice of maxim for which we are responsible. By viewing one's status this way, one can retain qualities associated with an innate propensity – i.e. it governs us from the beginning and its origin cannot be located in a causal chain – but since it is also being understood as a kind of *act*, it can nevertheless be characterized in terms attributable to freedom (Kant 1996: 74, 80; R 6: 25, 32).

The implications of this move are significant. Viewing this inaugural 'event' of moral failure as an intelligible act allows Kant to characterize it both as morally imputable *and* as cognizable by reason apart from any temporal event of explicit choosing. By characterizing the origin of

radical evil as an intelligible act Kant indicates that the human tendency to act on self-love must be understood as having a rational, not an empirical origin. In other words, our tendency to choose self-love over morality must be *conceived* as being ultimately grounded in an act of freedom that inverts the appropriate hierarchy – an act, therefore, which is not empirically enacted but rationally *posited* as the cause of all specific failures to live up to the moral law (Kant 1996: 70; R 6: 20). Finding ourselves always already having failed to perfectly follow reason, practical rationality demands that we posit a *cause* for this failure – a cause for which we can understand ourselves to be responsible. To call the source of radical evil an ‘act’, then, is somewhat misleading; its status as *intelligible* act means, rather, that we must treat this condition *as if* it were the consequence of an act so as to make our moral experience coherent. This is the sense of the prescriptive/descriptive distinction: characterizing oneself as having chosen one’s radically evil status is not a simple description of observed events; rather, it is to lay down as a rule *that* one will view oneself henceforth according to that interpretation – that one will view specific acts of self-love as arising from a global choice to prioritize them over morality. In taking this stance, then, one chooses to view what presents itself as a propensity as being an *act* – for it is only thus that one can make sense of one’s moral experience.

Although we cannot experience the origin of the distance between law and incentive, we can adopt a stance whereby the fact that sensuous incentives compete with moral incentives is reinterpreted as a general choice to make the latter conditional on their agreement with the former. Reason must *assume* that there was a first choice of this general misordering – a first choice whose very status as free requires it to be understood in terms of a *maxim*. Thus reason’s self-attribution is not simply of evil – of this or that moral failure – but of *radical* evil; namely, of a *principled* choice always to place incentives of self-love above the moral law even though such a principled choice was never ‘in fact’ made. Prior to this self-interpretation as evil, one interpreted one’s specific acts as ungoverned by such an overarching maxim and instead saw them as simply arising in response to each situation. By attributing radical evil to an act of freedom, however, one transforms an unchosen natural fact characterizing human beings into a morally fraught global *stance*. One sees each moral failure as the consequence of a fundamental choice to make morality conditional on self-love.⁶

This reading allows us to understand the relationship between the ‘supreme maxim’ governing the relationship between self-love and

morality, and the specific maxims governing the particular acts that arise from it. The 'supreme' maxim produced by attributing radical evil to a free act – namely, the stance according to which I take responsibility for prioritizing the moral law – must be operative in and through all particular maxims if they are to be considered moral at all. Otherwise, the fact that morality has overcome incentives of self-love in this or that particular case must be understood as *arbitrary*: having arisen purely from the particularity of the situation. The danger is – as Kant and many of his critics recognized – that the relationship between specific moral and sensuous incentives might be conceived as nothing more than a competition of strength.⁷ The notion of a global endorsement of one type of incentive over another provides Kant's solution to this problem. Attributing radical evil to oneself is choosing to view one's character as a whole on a moral register – ceasing, thereby, to see specific instances of moral failure as innocent cases of the sensuous outweighing the moral incentive. With the attribution of radical evil to one's freedom, one adopts a kind of meta-stance through which the general *relationship* of the incentive-types is viewed as itself being an object of choice. Since one recognizes that one has committed particular moral failures, this relationship is initially viewed as a general prioritization of self-love over morality – i.e. one sees oneself as radically evil.

Reason demands that we view our condition of moral failure as arising from an act of choice. Though we must attribute this condition to freedom, however, we are not thereby justified in concluding that a specific, temporally identifiable – i.e. empirical – choice to invert the hierarchy was in fact made at some point. Rather, such a self-interpretation is a necessary condition for the possibility of understanding oneself as responsible for a moral failure that one did *not* explicitly choose in this way. Indeed, in his definition of what a 'propensity' to evil entails, Kant claims that, although it *can* indeed be innate, yet it '*may* be represented as not being such: it can rather be thought of (if it is good) as *acquired*, or (if evil) as *brought* by the human being *upon* himself' (1996: 77; R 6: 29) – a point that captures the prescriptive nature of this self-understanding. Though the condition of moral imperfection is a natural one, reason transforms its status as such through a spontaneous act of reinterpretation – it comes to see this condition as having been brought upon itself. In doing so, an unchosen natural propensity to allow self-love to trump morality on this or that occasion becomes an act of freedom in which I see such events as indications of a *principled* failure to make self-love conditional on morality.

It is for this reason that Kant will claim that the propensity to radical evil is a universal tendency. In so far as humans do not have holy wills, we will always have displayed *some* moral failing – some instance in which reason and subjective incentive do not cohere. If we are to hold ourselves responsible for such specific failings, we must attribute to ourselves a principled choice to grant sensuous incentives pride of place. The only way in which the will can accomplish responsibility from a condition in which it finds itself having already chosen otherwise is if the overarching maxim by which it governs itself allows it to recognize conflict between law and incentive as something for which it is answerable. Thus the agent decides to take up the stance of self-responsibility by shifting from (i) a self-interpretation of her condition as a *natural* distance between incentive and law to (ii) a self-interpretation of her condition as a *failure* arising from her choice to permit this natural distance between incentive and law to produce immoral acts. Making oneself morally responsible for one's failings in this way means making oneself answerable *for one's very condition* as a non-holy will – viewing individual moral choices not as isolated events but as indications of the overarching stance one has adopted toward the role of morality in one's life. In so far as such an agent recognizes her specific moral failings *as* moral failings, she has adopted a stance whereby she sees the misordering of the incentives as attributable to her freedom.⁸ The agent engages in this kind of transformation of self-understanding when she imputes radical evil to her freedom – she takes as her maxim a general responsibility for her moral failures, regardless of what natural tendencies may be acting on her.

Every evil action must be so considered, *whenever we seek its rational origin*, as if the human being had fallen into it directly from the state of innocence. For whatever his previous behavior may have been, whatever the natural causes influencing him ... his action is yet free and not determined through any of these causes; hence the action can and must always be judged as an *original* exercise of his power of choice. (Kant 1996: 86-7; R 6: 41; emphasis mine)

By imputing radical evil to my freedom I refuse to understand acts conflicting with the moral law as isolated events arising from innocent natural influences. Rather, I consider myself to have fallen directly from innocence into guilt in the sense that I do not allow any exculpatory access to dimensions of my life that are not morally relevant. Henceforth all present temptations, natural dispositions, or past habits cannot excuse any moral failure that I commit. In this sense I make myself

responsible for what I am, answerable even for the things that I did not explicitly choose – the temptations working on me, the way in which I was raised. Though I did not cause these aspects of myself that work against the moral law, I am nevertheless able to take responsibility for the role they will play in my life. By imputing radical evil to my freedom, I choose to see the things that are simply working on me (i.e. sensuous incentives and self-love) as things for which I am morally answerable. This is not a causal responsibility – I continue to recognize that these things are mere ‘givens’ – but I can view my failure to place their demands second to those of morality as my fault. I am answerable for the role they play in relation to the moral law. Thus by interpreting oneself as radically evil, ‘I was born this way’ cannot excuse moral failure, since one has chosen to view such features of one’s being as secondary to the demand of the moral law. Through such a self-interpretation the agent integrates her condition of alienation into a narrative of moral autonomy by judging all actions in terms of whether they support or undermine the appropriate hierarchy of law and incentive.⁹

The agent who does *not* interpret himself as radically evil – i.e. who does not view the relationship established between the moral law and his incentives of self-love as attributable to his freedom – is still responsible for his moral failings in the sense that he can and must be held accountable for them by others. But because he has not yet adopted a stance whereby he holds *himself* accountable for such violations, he has not yet achieved the genuine responsibility of moral autonomy. In other words, he does not view his specific choices in terms of the role they will play in creating an appropriate relationship between incentive and law. Rather, he is heteronomous in the sense that each choice is experienced as an isolated event governed by the specifics of past events, natural impulses, or the occasional appeal of the moral law. Such an agent will sometimes be drawn by the moral law, other times not – but in each case the agent’s choice is governed by whatever incentive is strongest. Since this choice is not governed by an understanding of oneself as responsible for what incentives are permitted to be strongest – i.e. whether self-love is conditional on morality or vice versa – one has not yet achieved the stance necessary for autonomy, whereby one judges all of one’s acts in light of the freely chosen relationship that such acts will establish between law and incentive.

Although we cannot experience the origin of our moral imperfection, then, interpreting ourselves according to such an original ‘choice’ of evil

is a type of transcendental condition for the possibility of moral agency, since without this self-conception we could not understand ourselves as responsible for our moral failures. Without this act of self-attribution moral responsibility would be unable get off the ground – we could always simply blame failures to be moral on the natural incentives of self-love acting on us. Thus characterizing oneself as radically evil plays an essential role in Kant's moral philosophy because it makes sense of how the moral project can *begin* for agents. It articulates the move by which an agent begins to make herself autonomous – despite being born into a condition that resists it.

Interpreting the origin of humanity's propensity to evil as an intelligible act thus plays an analogous role to a postulate of practical reason, since it is an interpretation required for practical reason to function as it ought. As Kant says in *Critique of Practical Reason*, such postulates are a type of 'theoretical proposition, though one not demonstrable as such ... attached inseparably to an a priori unconditionally valid *practical law*' (1997a: 102; CPR 5: 122). What characterizes the postulates, in other words, is the fact that they are 'necessary conditions' for the proper functioning of practical reason (Kant 1997a: 110; CPR 5: 132) and they play this role not simply by providing reason with certain concepts but by *granting* these concepts a type of objective reality that they could not achieve through speculative reason alone:

These postulates are not theoretical dogmas but *presuppositions* having a necessarily practical reference and thus, although they do not indeed extend speculative cognition, they give objective reality to the ideas of speculative reason ... and justify its holding concepts even the possibility of which it could not otherwise presume to affirm. (Kant 1997a: 110; CPR 5: 132)

Practical reason *needs* to interpret the propensity to place self-interest above the moral law as the product of a free act – an act that cannot be *experienced* as such, but must be *assumed* as a necessary condition for the intelligibility and proper functioning of moral agency. By attributing radical evil to its freedom, reason in effect claims ownership of the condition in which it simply finds itself – transforming the fact of distance from the moral law into an act of principled alienation for which it is answerable.

Translating the Transcendental

The distinction between empirical and intelligible acts is therefore essential for reconciling the conflict inherent in the notion that one

chooses an *innate* disposition. The failure to maintain this distinction tends to be the norm, however. Thus Kant notes that the scriptural account takes original sin to be an initiating condition for subsequent deviations from the moral law. Unlike Kant's transcendental account – in which the self-ascribed *intelligible* act is taken to be a condition for the possibility of future evil acts because they cannot be rationally comprehended without it – the scriptural account makes an *empirical* act (eating the apple) the inaugurating condition for future sin simply because it predates it temporally (Kant 1996: 89; R 6: 43–4). The concept of 'original sin' essentially *translates* an intelligible act into a sensible act such that reason 'grants' a temporal location to what is in fact a transcendental condition of practical reason.

Nevertheless, humanity's desire for narrative translations of transcendental conditions – which Kant refers to as this 'weakness of ours' (Kant 1996: 88; R 6: 43¹⁰) – should not distract us from the fact that when we search for the origin of human evil, what we really seek is not the moment it began, but its 'inner *possibility*' (Kant 1996: 86; R 6: 41; emphasis mine). Namely, we are uncovering what we *must* take to be the case if we are to understand ourselves as responsible for our moral failings. One can note Kant's repeated claims that we must 'consider' and 'judge' it to be so whenever we seek the rational origin of evil. Reason is justified in assuming that the propensity to evil was chosen – in believing that it *must have been* chosen (one can note the 'as if' linguistic structures throughout) – because it is a necessary condition for *imputing* this alienation of the objective and subjective dimensions of the will.¹¹ Indeed, the same 'as if' formulations can be read throughout the *Groundwork* – Kant consistently speaks of the necessity of *thinking* of oneself or *calling* oneself free (Kant 1997b: 61; G 4: 457).

Putting on the New Man

If reason is to understand itself as responsible for the condition in which it simply finds itself, it demands a fundamental shift in self-understanding: the moral agent cannot think of her condition as an unchosen natural state but must interpret herself as *responsible* for that condition. Just as *pure* reason must apply interpretative categories in order to make its experience of physical nature lawlike and coherent, so too must *practical* reason apply certain interpretative categories to make the exercise of its freedom lawlike and coherent.

Unlike the other postulates of practical reason, however, the postulate of radical evil is not merely a *theoretical* proposition granted an

objective reality that it could not achieve through speculative reason alone. Rather, the assumption of one's 'radical evil' is a self-interpretation that serves to *enact* the very condition that it assumes. In other words, by understanding myself as the origin of my moral failings – which I do when I attribute the misordering of my incentives to a principled act of freedom – I first *realize* myself as a responsible being. Understanding radical evil as an intelligible act is not simply information that practical reason needs in order to function; it is, rather, a *performative* self-relation in which I first *take over* and *impute* responsibility to myself for my condition.¹² To see specific choices of self-love over morality as a principled failure of freedom – not as individual natural events – is therefore essential to the project of becoming genuinely moral, as Kant's distinction between moral dogmatics and moral discipline makes clear:

The thesis of innate evil is of no use in moral *dogmatics*, for the precepts of the latter would include the very same duties, and retain the same force, whether there is in us an innate propensity to transgression or not. In moral *discipline*, however, the thesis means more, yet not more than this: We cannot start out in the ethical training of our connatural moral predisposition to the good with an innocence which is natural to us but must rather begin from the presupposition of a depravity of our power of choice in adopting maxims contrary to the original ethical predisposition. (1996: 94; R 6: 50–1)

We must begin with the *presupposition* of our own fault – for it is only thus that we can start on the road to self-ownership. What makes the self-imputation of radical evil a unique type of postulate, then, is the fact that through the interpretation of oneself as morally responsible for one's failures, this objective reality – i.e. responsibility for one's moral status – is not merely posited but is first *brought into being*. By seeing specific moral failures as indications of freedom's choice to make morality conditional on self-love, one places oneself in the stance that would allow for the reversal of the hierarchy. One sees the relationship between moral and sensuous incentives as something to be *chosen* – not simply lived.

This inauguration into the stance of responsibility is related to Kant's notion of 'putting on the new man', an event which 'can come about only through a kind of rebirth, as it were a new creation ... and a change of heart' (1996: 92; R 6: 47). To become fully responsive to the non-negotiable nature of our moral obligations and their absolute

priority, Kant argues, ‘a revolution is necessary in the mode of thought’ (1996: 92; R 6: 48) – i.e. the way in which we *understand* ourselves. ‘Putting on the new man’ is such a revolution in disposition: the agent engages in a ‘single and unalterable decision’ that ‘reverses the supreme ground of his maxims’ and thereby becomes a ‘subject receptive to the good’ (Kant 1996: 92; R 6: 48). Such a reversal does not entail that one will henceforth be completely moral; this revolution in thought still requires a ‘gradual reformation in the mode of sense’ – i.e. one must work to manifest this different self-understanding in action by struggling ‘upon the road of endless progress toward holiness’ (Kant 1996: 91; R 6: 47). Nevertheless, this transformation into responsible selfhood – setting out on the road toward holiness – is first entered into with the self-imputation of guilt as a *principled* way of answering up to the condition in which we find ourselves.

‘Putting on the new man’ and the self-attribution of radical evil cannot be equated, then, since the former describes the moral transformation from evil person to good person, whereas the latter merely involves seeing oneself as an evil person. In the former case, the agent decisively reverses the hierarchy between self-love and morality such that henceforth all actions rooted in incentives of self-love are made conditional on their agreement with morality. Such a person continues to struggle with actually subordinating this or that recalcitrant incentive to morality, but he has ‘put on the new man’ in the sense that the supreme ground of his maxims has been reversed from prioritizing self-love to prioritizing morality. Such a revolution is only possible, however, if one has already come to see oneself through the lens of such a freely chosen hierarchy. In other words, such a reversal in priority depends on having adopted a principled stance for understanding one’s moral condition in the first place – as one does when one views specific moral failures as arising from a free choice to make morality dependent on self-love. Conceiving of oneself as *globally* responsible for which incentives will be incorporated into one’s maxims in this way is to take a principled stance on the relationship between incentives of self-love and incentives of reason, rather than allowing this relationship to be determined by the particularity of the situation – the condition that characterizes one prior to the self-imputation of radical evil. Because this imputation is an intelligible act – not an empirical one – taking responsibility here does not mean being the *cause* of the self-love that has interfered with the moral law, but being *answerable* for it.¹³ By taking myself to be *responsible* for the fact that I am in a condition in which incentives of self-love have priority over those of duty – calling myself ‘evil’ – I have in effect made the first step

towards reversing those terms – a reversal that is only fully accomplished when I ‘put on the new man’ by making self-love dependent on reason.

Humiliated Freedom

We have seen, then, that when faced with the lack of coincidence between the objective obligating dimension and the subjective responsive dimension of the will, practical reason’s response is to *find a way* to view this condition as something for which we are responsible. We must understand ourselves not only as responsible for the lack of coincidence between moral obligation and moral responsibility, however; we must also overcome the tendency to avoid recognizing our contribution to this moral distance. The question, then, is precisely how the agent is motivated to attribute a failure of responsibility to herself when she is not yet responsible (Kant 1996: 92; R 6: 47). Unless he can answer this, Kant will be unable to explain how moral agents are *motivated* to take up the prescriptive injunction to own up to responsibility in the way we have been articulating. In other words, he will fall prey to a version of Sidgwick’s objection that Kantian ethics makes moral responsibility for bad things impossible, only here it would make moral responsibility entirely optional, since the status of ‘morally responsible’ would depend entirely on the choice to call oneself radically evil – to attribute the gap between reason and incentive to an act of freedom. We must answer, then, why the agent would be motivated to take such an interpretative stance. Why would one not persist in the pre-moral simplicity of childhood, clinging to the sense that incentives of self-love are simple givens for which one cannot be responsible?

It is in response to this requirement, I believe, that even when decrying the ‘universal’ status of radical evil, Kant notes that this ‘innate’ tendency can nevertheless only be discerned at ‘the first manifestation of the exercise of freedom’ (1996: 84; R 6: 38). This ‘innate propensity’ to place self-love above reason is not evident in children because prior to the ‘age of reason’, one is not claimed by the dictates of reason. One is therefore incapable of the reversed priority of incentives characterizing radical evil. It is only when the youth first exercises his freedom – i.e. only with the appearance of rational action – that this supposedly ‘innate disposition’ comes to be recognized as such. This is not (only) because teens are immoral; rather, instances in which inclinations of self-love are preferred to the moral law are only possible – and thus imputable to oneself as such – once one is capable of rational action (i.e. once one is capable of succeeding or failing at acting on reasons). Only when one is mature enough to engage in free acts can one recognize that one *ought* to take responsibility for this capacity.

Once having reached the age of reason, however, natural self-love tends to be taken for an overriding reason for action – a result of what Kant refers to as a kind of ‘dishonesty, by which we throw dust in our own eyes and which hinders the establishment in us of a genuine moral disposition’ (1996: 84; R 6: 38).¹⁴ Such ‘dust-throwing’ involves clinging to self-interpretations that promote the belief that one need not take responsibility for one’s moral condition, especially if one typically follows the letter of the law or has been lucky enough to escape the negative consequences of one’s moral failures (Kant 1996: 84–5; R 6: 37–9). Even after the appearance of reason and its claims we desire to remain in childhood and its innocent immersion in self-love. With the advent of reason, however, that innocence is lost – simple unthinking immersion in self-love becomes a deliberate effort to believe that self-love provides overriding reasons to act. Despite the tendency to deny one’s changed condition, however, Kant wants to demonstrate that the moral law breaks through to claim one regardless. Something about the youth’s capacity for free action must motivate him to follow Kant’s prescriptive injunction to ascribe radical evil to his freedom. But how does the exercise of freedom itself motivate us to recognize that we are failing to take responsibility for rationality, as we know we ought to do?

The exercise of freedom is normatively structured according to maxims that establish ends and the means necessary to accomplish them. The youth’s ability to engage in this kind of practical rationality therefore involves an implicit endorsement of the faculties that allow such principled action. By acting freely, one commits oneself to the constraints of practical rationality that make such freedom possible. Nevertheless, one is capable of making use of practical rationality in the pursuit of one’s ends while exempting oneself from the constraints that characterize its implementation. In such a performative contradiction, the subjective inclination *not* to prioritize reason contradicts the very rationality that is at work in taking one’s projects of self-interest as ends that provide reasons to act.¹⁵ In such a condition one ignores the lawfulness implicit in all acts of practical rationality and chooses, instead, to endorse the heteronomy that undermines it.¹⁶ Or, as Kant says, one exists in a condition where one’s rational humanity – and not one’s responsible personality – dominates. The youth’s capacity for rational action means, then, that he has entered the normative space of humanity. But prior to taking responsibility for this rationality – which occurs when he attributes his general condition of moral imperfection to an act of freedom – he has not yet achieved the level of genuine moral personhood.

By describing how rational agents can recognize that they are always obligated by the moral law, Kant is able to motivate the prescriptive demand that they attribute radical evil to their freedom and thereby make themselves answerable for failures to meet this obligation. By showing how the moral law breaks through, in other words, Kant can respond to Sidgwick's objection, since he can show that, despite being a *choice*, it is a highly motivated choice to which we feel compelled to respond. It is in this sense that we must understand Kant's empirical assertion about the universality of the experience of moral failure. This claim is a type of exhortation to the reader – a call to recognize the implications of her own rationality. It asks one to acknowledge that between the sleep of childhood and the clairvoyance of full-blown autonomy, there is a phase of human life during which one tacitly feels the pull of the moral law – and so could in principle acknowledge its supremacy, despite one's desire to persist in the narcissism of childhood. This is a phase during which one acts on incentives of self-love despite the sense that one must give reason its due. By recognizing *that* one is failing in such cases, one is motivated to attribute entry into this condition of moral failure to oneself as a free choice.¹⁷ Only by feeling the failure *as* a failure would one take up a stance of self-ownership whereby one claims the failure as something for which one chooses to make oneself answerable.

The movement towards full moral agency is motivated, then, by acknowledgement of the implicit constraints of practical reason. Kant argues that it is the representation of the moral law operative in such an acknowledgement – i.e. the representation of the agent's capacity to be truly free – that *humiliates* the agent who has subordinated that freedom to the very inclinations that thwart it. Agents recognize simultaneously that they are capable of acting rationally and yet failing to consistently do so. This experience produces a 'painful' interruption of the choice to act on the incentives of self-love. Such a painful event – the humiliation of realizing that one is subordinating freedom to the merely natural inclinations – motivates the negative self-ascription found in the notion of radical evil.¹⁸ Ascribing radical evil to one's freedom captures the experience of this painful and humiliating event. By interpreting oneself in terms of an innate propensity towards a kind of global evil, one takes ownership of the normative space into which one has entered through the exercise of practical rationality. Such a self-imputation is painful because one characterizes oneself as a failure not simply at this or that exercise of rationality, but at being rational *in general*, since one has failed to take rationality as the ground for

one's actions.¹⁹ Kant's characterization of us as '*evil*', then, plays an important rhetorical role because it captures the sense that the human being is failing to be his 'proper self', a self that he becomes only when he

claims for himself a will which lets nothing be put to his account that belongs merely to his desires and inclinations, and on the contrary thinks as possible by means of it – indeed as necessary – actions that can be done only by disregarding all desires and sensible incitements. (1997b: 61; G 4: 457)

The profoundly negative tenor of imputing 'radical evil' to oneself is in this respect no accident; claiming responsibility involves calling oneself out as a moral failure, and the painful dimension of this calling out is essential for interrupting the pull of other incentives (Kant 1997a: 63; CPR 5: 72–3). It tears away comforting delusions about the negotiable character of the moral law by bringing the agent face to face with the very freedom that is being thwarted by her failure to give it its proper due. By attributing radical evil to herself, the agent acknowledges the general capacity to fail to act on reasons by recognizing that she has been acting on natural impulses and external motivations. This profoundly negative experience is not mere Pietist self-loathing; it is, rather, a stance necessary for the self-transformation that places one on the road to autonomy.²⁰ Without this presupposition we would be required to consider ourselves as beasts or gods: either heteronomously constrained by nothing but natural dispositions for which we could not be held accountable, or holy wills to which the notions of obligation and accountability do not apply. Either case would unduly simplify the complex dynamic of human agency – the fact that we understand ourselves *as* failing to meet our moral obligations in the very moment that we are first capable of freely meeting them.²¹

Respect

Full moral autonomy demands that an agent take explicit responsibility for the lawfulness to which she has been implicitly committed from the moment she began to exercise her freedom. By adopting an interpretative stance in which the misordering of the incentives is recognized *as* a misordering for which one can be responsible, one refuses to understand oneself as simply determined by the given conditions. In so doing one takes the first step into the arena of moral responsibility. However, Kant must also account for the will's *positive* stance towards its obligating dimension – the manner in which the subject *responds* to its obligations once competing claims have been 'humiliated'.

Thus the reconceptualizing that occurs in the self-imputation of radical evil must involve not only a subjective appropriation of one's *failures*; it must also provide a subjective incentive to enact the responsibility that one has claimed in that act of appropriation. Kant must explain *why* we are motivated to 'put on the new man'. For Kant, this incentive occurs in the experience of *respect*: 'If something represented as a *determining ground of our will* humiliates us in our self-consciousness, it awakes *respect* for itself insofar as it is positive and a determining ground' (Kant 1997a: 64; CPR 5: 74). In respect we are moved by the demands of the moral law. Respect names the moral law's 'effect on feeling', thereby indicating how the moral law itself – the *objective* determining ground of the will that is claiming us – 'awakes' in us a corresponding '*subjective* determining ground – that is, an incentive' (Kant 1997a: 64–5; CPR 5: 74–6).²² With the notion of respect Kant answers how the moral law itself *creates* a subjective inclination to be responsive to its claims: in respect, practical reason 'effects a feeling conducive to the influence of the law upon the will' (Kant 1997a: 65; CPR 5: 75). However, the moral law does not simply make use of a pre-existing feeling but *initiates* or *evokes* this feeling in the subject – a feeling which is, therefore, not pathologically but *practically* effected: 'And so respect for the law is not the incentive to morality; it is morality itself subjectively considered as an incentive inasmuch as pure practical reason ... supplies authority to the law' (Kant 1997a: 65; CPR 5: 76). Respect is the incentive to give morality pride of place, making incentives of self-love conditional on their conformity to the law.

Through the representation of the moral law – through practical reason's representation of the agent as failing to meet the obligations by which she is bound – the first act of responsiveness to this obligation is initiated. Through this self-understanding one inaugurates the stance of self-ownership – thus we can note Kant's shift from characterizing respect as an 'incentive to morality' to 'morality itself'. The subjective and objective determining grounds of the will are being brought into unity through this *creation* of a responsible stance by obligation itself. Thus respect 'supplies authority to the law' – an interesting formulation that demonstrates the essential distinction between *knowing* the obligation of the law and feeling oneself *bound* by it.²³ Since responsibility demands not only that one *is* obligated but that one experience oneself *as* obligated, Kant introduces the notion of respect to demonstrate how the obligations of reason are first subjectively *experienced* as having such a motivating claim. Only in light of such an experience will the agent be motivated to transform the hierarchy of incentives such that incentives of self-love are made conditional on their agreement with the moral law.

Conclusion

The imputation of radical evil – whereby the subject calls herself out as a moral failure and takes a first step towards reorienting her priorities through the experience of respect to which this self-imputation gives rise – is a kind of inaugurating instance of responsibility. It is a birth into moral selfhood that cannot itself be grounded in anything other than the unconditioned nature of freedom itself, despite our seemingly inevitable need to inquire into that which came before (Kant 1996: 74; R 6: 25). By interpreting itself as globally evil the self makes itself answerable for a condition for which it is *not* responsible in so far as it simply finds itself torn between self-love and the moral law from the minute it first makes use of reason to pursue its projects. Rather than wallowing in heteronomy – viewing each instance of moral failure as the natural result of morality’s competition with stronger incentives of self-love – the autonomous agent chooses to see such instances as indications of a principled failure of freedom to establish the correct hierarchy of incentives. The self-imputation of radical evil thus involves a kind of taking responsibility *for* responsibility – the first step in the ‘establishment in us of a genuine moral disposition’ (Kant 1996: 84; R 6: 38). To ‘assume’ one’s status as radically evil, then, does not mean to posit it without theoretical justification – it means to own it as one’s own.

Contrary to a certain tendency in Kant scholarship, then, the notion of ‘radical evil’ cannot be dismissed as religious baggage. The interpretation of oneself as the origin of radical evil is not a theological holdover or an anthropological claim but a transcendental condition for practical reason’s understanding of itself as morally responsible. Thus my prescriptive reading of radical evil demonstrates that Kant’s work in the *Religion* is in line with his earlier works. Despite being couched in religious terminology, Kant is breaking with the backward-looking notion of ‘original sin’, whereby we attempt to make sense of being born evil by offering a metaphysical account of why we deserve it. Rather, his account demands a forward-looking appropriation of one’s past in light of the future self that one struggles to be. Indeed, in reading Kant this way we can solve one of the traditional criticisms of Kantian ethics – namely, that it has no developmental account, but merely *presupposes* autonomous agents.²⁴ The notion of radical evil is Kant’s solution to this developmental issue – a solution that rests on the idea that a transformation in self-understanding starts one on the road to autonomy. Despite the negative tenor of understanding oneself as ‘radically evil’, then, what is being articulated is *self-respect*: a stance in which one makes oneself answerable for one’s failures and in so doing first constitutes oneself as a responsible being.

Notes

- 1 See O'Connor (1985: 288–302), Michalson (1990), and Bernstein (2002). Bernstein argues that, despite Kant's rhetoric in the *Religion*, his concept of radical evil 'turns out to be little more than a way of designating the tendency (propensity) of human beings to disobey the moral law' (2002: 43). And as O'Connor notes, 'There is something odd about Kant's whole discussion of evil, for even if we accepted Kant's contradictory notion of an evil disposition which is freely chosen outside of time, we would by his own admission gain nothing in the way of understanding insofar as the origin of the perversion of the will would remain inscrutable to us.' (1985: 239). The religious interpretation suggests itself when we note that his account of radical evil is quite similar to the traditional Augustinian understanding of original sin, according to which original sin is an innate inheritance from Adam for which we are nevertheless guilty. Though Kant rejects Augustine's notion of *inheritance* (see his discussion in *Religion Within the Boundaries of Mere Reason* (= *R* in the text; 1996: 86; *R* 6: 40, footnote), he nevertheless adopts an Augustinian stance in so far as he claims that all human beings are culpable for their innate propensity to evil. See Quinn (1990) on this point. Michalson similarly argues that Kant's talk about innate evil requires him, ultimately, to come to rely on the role of divine salvation in a way that contradicts – or at least seriously complicates – his notion of moral autonomy. See Frierson (2010) and Hare (1996) for accounts that read the *Religion* as endorsing humanity's dependence on God's grace for overcoming this innate evil. Allen Wood argues, on the other hand, that the notion of radical evil should be cashed out simply in terms of Kant's anthropological claims about humanity's natural tendency towards social comparison and the antagonism that this produces (1999: especially 288–9). See also Sullivan (1995).
- 2 This optimism is evident in his works on history, where he suggests that we should be hopeful about humanity's capacity for moral progress. Kant is far from naïve on this point, however, as Susan Neiman shows (2002). See also Morgan (2005: 106).
- 3 Morgan (2005: 65). Morgan argues that the evil will allows itself to believe that moral (internal) freedom is no different from external freedom or lack of constraint – even though the moral agent knows what the moral law demands of her. 'The only way the will can have been motivated by its spontaneity to subject itself to causality in the form of the choice of self-love is through its representation to itself of freedom as the unlimited indulgence of all its whims' (Morgan 2005: 86–7). Henry Allison argues that Kant's position arises naturally out of his rigorism – the view that acts and agents are either good or evil, with no middle ground between them (Kant 1996: 71–2; *R* 6: 22). In the absence of a propensity to good, rigorism demands that one *must* be characterized by a propensity to evil. Thus to show why a propensity for evil must be the case for all human beings, Kant merely has to show why an agent characterized by a propensity for good is impossible. He can accomplish this, Allison argues, because for such a being 'there could be no possibility of temptation and no thought of the law as constraining, which, in turn, means that the law would not be viewed as an imperative and its requirements as duties'. Since humanity's desire for happiness means that there is always the possibility of temptation, he argues, moral requirements will always be experienced as duties and radical evil is *a priori* true of all human beings (Allison 2001: 609).
- 4 Korsgaard examines this connection in detail in *Creating the Kingdom of Ends* (and elsewhere), noting that 'since the will is practical reason, it cannot be conceived as acting and choosing for no reason. Since reasons are derived from principles, the free will must have its own principle' (1996a: 163).
- 5 This interpretation is indebted to Tengelyi (2004), Allison (1990) and Ricoeur (2007).

- 6 I take Allison's account of the role of rigorism in the missing formal proof to be right then, except that my interpretation emphasizes the *prescriptive* role that such rigorism plays. Thus morally indifferent actions are not possible for a *moral* agent because their very 'indifference' would presuppose that the agent had not yet adopted the stance of radical responsibility through the self-attribution of evil: 'A morally indifferent action (*adiaphoron morale*) would be one that merely follows upon the laws of nature, and hence stands in no relation at all to the moral law as law of freedom' (Kant 1996: 72; R 6: 23, footnote). See also Kant's note that rigorism is appropriately applied when we are considering what we *ought* to be – it is not meant to be an adequate description of what human beings are as they *appear* (1996: 73; R 6: 25, footnote).
- 7 Or, as Dean Moyar puts it, 'the textbook "rigorist" Kant obsessed with the battlefield of motivational forces' (2003: 196). See also Sussman (2001).
- 8 This prescriptive emphasis distinguishes my account from that of David Sussman, who argues that such a global disposition refers to nothing more than the shape of one's moral life as a whole (2005: 173). Similarly, Korsgaard's characterization of the intelligible act as a kind of thought-experiment in which the will decides what principle will govern it prior to its entry into the world (1996a: 164–7) – an idea taken up by Morgan (2005: 76–8) – is right in a sense, yet fails to accommodate fully the performative, prescriptive dimension that I take to be essential.
- 9 Reading Kant this way also allows us to avoid characterizing this original choice of evil as occurring from a sheer liberty of indifference (see Quinn 1990: 231) – the reason being, of course, that there was *in fact* no such choice at all.
- 10 See also *Groundwork of the Metaphysics of Morals*: the human being 'is very much inclined to expect behind the objects of the senses something else invisible and active of itself – but it spoils this again by quickly making this invisible something sensible in turn, that is, wanting to make it an object of intuition, so that it does not thereby become any the wiser' (Kant 1997b: 57; G 4: 452).
- 11 Evgenia Cherkasova reads such formulations, on the contrary, as evidence that Kant is attempting to 'express the connection between the knowable and the unknowable'. She concludes that these are fictions, however – useful for ethical inquiry but not constitutive of the ethical stance itself (2005: 582–3).
- 12 As noted above, Allison also suggests that radical evil can be understood as a type of postulate of pure practical reason, but my account differs insofar as it is aimed at demonstrating how this self-imputation of guilt *inaugurates* moral agency. See Allison 1990: 146–61.
- 13 This distinction is important because it leaves room for understanding the self as having a dignity irreducible to the natural order. Reading radical evil in terms of the moral phenomenology of developing responsibility, then, does not commit me to a deflationist account of dignity. Rather, it is in keeping with Kant's view that understanding oneself as free means understanding oneself in terms of the dignity of noumenal – not purely empirical – being. Note, however, that this self-understanding and the noumenal/phenomenal distinction operative in it is produced by *practical* – not theoretical – reason.
- 14 Henry Allison notes the importance of self-deception for a Kantian account of evil in his article 'Reflections on the Banality of (Radical) Evil: A Kantian Analysis' (Allison 1996: 169–82).
- 15 These structures are discussed throughout Korsgaard's work, including 'Kant's Formula of Humanity' (Korsgaard 1996a: 106–33, and 1996b). Donald Regan (2002) challenges the idea present in her work that value is dependent on such rational endorsement. David Sussman (2003) shows that this criticism is justified only if the claim is that practical reason's willing *creates* goods, rather than the more

- modest claim that rational nature is a condition for the possibility of their *being* goods for an agent.
- 16 Yirmiyahu Yovel points out that this is facilitated by the fact that the dimension of will responsible for particular acts of choosing – or *Willkür* – makes a kind of *claim* to self-determination in its choosing, but it is only when *Willkür* recognizes the universality of *Wille* as necessary for its realization that the will achieves genuine self-determination (1998: 289–91).
 - 17 Kant addresses this developmental issue somewhat in the text ‘Conjectural Beginning of Human History’ where he discusses the qualities necessary for humanity as opposed to mere animality. These include such things as the capacity for comparisons, an understanding of temporality (especially futurity), and the creation of cultural needs and morally relevant emotions such as anxiety, shame, care. In his closing remarks to that text we can note Kant’s exhortation to the reader to be content with providence ‘in part in order to not, by placing the blame on fate, lose sight of our own fault, which may perhaps be the only cause of all these ills, and fail to seek help against them in self-improvement’ (2006: 24–36; 8: 121).
 - 18 A complete account would demand an articulation of the motivating role that others play in taking responsibility for one’s freedom – a position worked out first by Fichte (2000) but since pursued by thinkers such as Emmanuel Levinas, many feminist theorists and (more recently) Stephen Darwall. However, despite the role that the second-person claim may play in motivating the agent to take responsibility, the structure of self-attribution required to do so remains the same. Taking responsibility demands that one recognize and answer for one’s *own* failures – though others help me acknowledge this demand. This is not to deny, however, that the normative space itself – the arena in which the agent exercises her agency (or indeed, this agency itself) – is social through and through. The degree to which our rationality depends on others, however, is regrettably beyond the scope of this paper.
 - 19 For a discussion of the manner in which practical reason’s goal is ultimately itself, see Yovel (1998: 273).
 - 20 Despite worries about the self-ascription of ‘radical evil’ being the stance of a masochistic superego, one can recognize that this kind of self-interpretation is in the service of a higher kind of self-love: i.e. the agent rejects aspects of herself in the service of becoming her *genuine* self. Korsgaard makes note of a similar point when she claims ‘the moral principle is not a principle which stands over and above the tendency to self-love, checking and correcting and limiting it. Instead, it is a kind of definition of it, a filling out who the *me* is that is the object of my self-love’ (1998: 54).
 - 21 This is the manner in which the ‘species’ claims in the ‘Religion’ must be read: as articulating a *subjectively* necessary self-attribution of a condition in which all humanity finds itself. Thus he notes that “‘He is evil *by nature*” simply means that being evil applies to him considered in his species; not that this quality may be inferred from the concept of his species ([i.e.] from the concept of a human being in general, for then the quality would be necessary), but rather that, according to the cognition we have of the human being through experience, he cannot be judged otherwise, in other words, *we may presuppose evil as subjectively necessary* in every human being, even the best.’ (1996: 80; R 6: 32, emphasis mine)
 - 22 For a more detailed account of the subjective dimension of the moral law’s obligating force, see Guevara 2000.
 - 23 Andrews Reath (1989) argues that the issue is not merely a battle for dominance between two psychic forces, but a struggle for *legitimacy*. See also Yovel (1998: 287–9) on this point.
 - 24 Barbara Herman (1998) discusses this issue. See also Sussman (2005) on this point.

References

- Allison, H. (1990) *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.
- (1996) *Idealism and Freedom*. Cambridge: Cambridge University Press.
- (2001) 'Ethics, Evil, and Anthropology in Kant: Remarks on Allen Wood's *Kant's Ethical Thought*'. *Ethics*, 111, 594–613.
- Bernstein, R. A. (2002) *Radical Evil: A Philosophical Interrogation*. Malden, MA: Blackwell.
- Cherkasova, E. (2005) 'On the Boundary of Intelligibility: Kant's Conception of Radical Evil and the Limits of Ethical Discourse'. *Review of Metaphysics*, 58, 571–84.
- Fichte, J. G. (2000) *Foundations of Natural Right*. Ed. F. Neuhouser, trans. M. Baur. Cambridge: Cambridge University Press.
- Frierson, P. (2010) 'Kantian Moral Pessimism'. In S. Anderson-Gold and P. Muchnik (eds), *Kant's Anatomy of Evil* (Cambridge: Cambridge University Press), pp. 33–56.
- Guevara, D. (2000) *Kant's Theory of Moral Motivation*. Boulder, CO: Westview Press.
- Hare, J. (1996) *The Moral Gap*. Oxford: Oxford University Press.
- Herman, B. (1998) 'Training to Autonomy: Kant and the Question of Moral Education'. In A. O. Rorty (ed.), *Philosophers on Education* (New York: Routledge), pp. 255–72.
- Kant, I. (1996) 'Religion Within the Boundaries of Mere Reason'. In A. W. Wood (ed.), *Religion and Rational Theology* (Cambridge: Cambridge University Press), pp. 41–215, Abbreviated R.
- (1997a) *Critique of Practical Reason*. Ed. and trans. Mary Gregor. Cambridge: Cambridge University Press. Abbreviated CPR.
- (1997b) *Groundwork of the Metaphysics of Morals*. Ed. and trans. Mary Gregor. Cambridge: Cambridge University Press. Abbreviated G.
- (2006) 'Conjectural Beginning of Human History'. In P. Kleingeld (ed.) and D. L. Colclasure (trans.), *Toward Perpetual Peace and Other Writings on Politics, Peace, and History* (New Haven, CT: Yale University Press), pp. 24–36.
- Korsgaard, C. (1996a) *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.
- (1996b) *The Sources of Normativity*. Cambridge: Cambridge University Press.
- (1998) 'Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind'. *Ethics*, 109, 49–66.
- Michalson, G. E., Jr. (1990) *Fallen Freedom*. Cambridge: Cambridge University Press.
- Morgan, S. (2005) 'The Missing Formal Proof of Humanity's Radical Evil in Kant's "Religion"'. *Philosophical Review*, 114, 63–114.
- Moyar, D. (2003) 'David Sussman, the Idea of Humanity: Anthropology and Anthroponomy in Kant's Ethics'. *Ethics*, 114, 196–9.
- Neiman, S. (2002) *Evil in Modern Thought*. Princeton, NJ: Princeton University Press.
- O'Connor, D. (1985) 'Good and Evil Disposition'. *Kant-Studien*, 76, 288–302.
- Quinn, P. L. (1990) 'Does Anxiety Explain Original Sin?' *Noûs*, 24 (APA Central Division Meeting), 227–44.
- Reath, A. (1989) 'Kant's Theory of Moral Sensibility: Respect for the Moral Law and the Influence of Inclination'. *Kant-Studien*, 80, 284–302.
- Regan, D. H. (2002) 'The Value of Rational Nature'. *Ethics*, 112, 267–91.
- Ricoeur, P. (2007) *The Conflict of Interpretations: Essays in Hermeneutics*. Ed. Don Ihde. Evanston, IL: Northwestern University Press.
- Sullivan, R. J. (1995) 'The Influence of Kant's Anthropology on his Moral Theory'. *Review of Metaphysics*, 49, 77–94.
- Sussman, D. (2001) *The Idea of Humanity: Anthropology and Anthroponomy in Kant's Ethics*. New York: Routledge.

- (2003) 'The Authority of Humanity'. *Ethics*, 113, 350–66.
- (2005) 'Perversity of the Heart'. *Philosophical Review*, 114, 153–77.
- Tengelyi, L. (2004) *The Wild Region in Life History*. Trans. G. Kállay and L. Tengelyi. Evanston, IL: Northwestern University Press.
- Wood, A. (1999) *Kant's Ethical Thought*. New York: Cambridge University Press.
- Yovel, Y. (1998) 'Kant's Practical Reason as Will: Interest, Recognition, Judgment and Choice'. *Review of Metaphysics*, 52, 267–94.