Kantian Review, 22, 3, 395–419 © Kantian Review, 2017
doi:10.1017/S1369415417000140

Can Kant's Theory of Radical Evil Be Saved?

ZACHARY J. GOLDBERG
Ludwig-Maximilians-Universität Munich
Email: zachary.goldberg@lmu.de

Abstract

In this article, I assess three contemporary criticisms levelled at Kant's theory of evil in order to evaluate whether his theory can be saved. Critics argue that Kant does not adequately distinguish between evil and mundane wrongdoing, making his use of the term 'evil' emotional hyperbole; by defining evil as the subordination of the moral law to self-love his analysis is seemingly overly simplistic and empirically false; and by focusing solely on the moral character of the perpetrator of evil, Kant's theory apparently ignores the most salient aspect of evil – the suffering of victims. While I will not claim that Kant provides us with a fully adequate theory of evil, I respond to each of these criticisms and conclude that Kant's theory can still provide significant insight into both the nature of evil and the moral psychology of perpetrators of evil.

Keywords: radical evil, mundane wrongdoing, root of evil, empirical harm, moral character

1. Introduction

In much of the recent philosophical literature concerning the nature of moral evil, it is common for scholars to reject or dismiss Kant's conception of radical evil.[1] Kant is one of the first philosophers to offer a decidedly secular theory of evil, and this historical precedence makes grappling with his account a natural starting point for theorists promoting new ideas about evil. In this article, I shall assess the main criticisms levelled at Kant in order to evaluate whether his theory can be saved. While I will not claim that Kant provides us with a fully adequate theory of evil, I respond to each of the predominant criticisms and conclude that his theory can still provide significant insight into both the nature of evil and the moral psychology of perpetrators of evil.

I will address three persistent criticisms directed at Kant's theory. First, scholars of evil are in agreement that, if the term 'evil' is to do any

explanatory work, it must refer to a category of action or moral opprobrium that is distinct from mere wrongdoing. Some critics argue that Kant fails to draw such a distinction and uncritically lumps the most severe types of evil together with fairly mundane acts of wrongdoing.[2] As a result, he cannot sufficiently account for the morally worst sorts of acts or persons. Second, some critics assert that Kant's reduction of evil to a single cause – the prioritization of self-love over the moral law – is overly simplistic and ignores the plausible notion that evil has many different roots such as cruelty, hatred, obedience, revenge, greed, thoughtlessness and religious faith, to name a few.[3] These critics conclude that Kant's reductive definition provides a much too narrow account of evil, which excludes other motives and intentions that reliably result in evil. Third, many scholars reject Kant's theory of evil because it seemingly fails to pay heed to the most salient aspect of an evil act – the actual empirical harm suffered by victims.[4] They charge that by defining evil solely in terms of the perpetrator's quality of will, he ignores the suffering that is crucial to any compelling theory of evil.

The goal of the present examination is to make the case for revisiting Kant's theory of radical evil for scholars of contemporary theories of evil. In addressing each criticism, my analysis will show the continuing relevance of Kant's theory to contemporary approaches to moral evil. I do not believe that the current investigation remedies all of the theory's philosophical problems nor inoculates the account against future criticisms. However, my argument demonstrates that Kant's theory can still help us understand the nature of moral evil and the moral psychology of perpetrators of evil.

2. Kant's Theory

In *Religion within the Boundaries of Mere Reason*[5] Kant presents his theory of radical evil, which is based on a complex moral psychology. For Kant, the incentive to adopt the moral law as one's maxim arises out of an original predisposition (*Anlage*) to good.[6] At the same time, we also have a natural propensity (*Hang*) to prioritize our sensuous inclinations over incentives of the moral law. The way in which the individual prioritizes moral and non-moral incentives defines her moral character as either good or evil. The subjective adoption of maxims that contain the objectively valid moral law as their end is constitutive of a good character. The subjective adoption of maxims containing some end other than the moral law is constitutive of an evil one. Evil is 'radical' in the sense of being rooted in the individual when prioritizing sensuous over moral incentives becomes the ground of an individual's maxim-making resulting in the

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

regular and continuous subordination of the moral law to sensuous incentives.[7] Kant refers to this regularity as one's disposition (*Gesinnung*) – as the subjective ground of all maxims. We might think of a disposition as an individual's frame of mind if by that we understand the basis – the underlying set of beliefs, interests and intentions – on which we can expect that individual to act.[8] Since the propensity to evil corrupts the subjective foundation upon which an individual adopts maxims, it is the source of every evil deed and therefore radical.[9] Despite evil's radical quality, Kant maintains that individuals at all times remain free and morally obligated to amend the ordering of their incentives.[10]

Although Kant's use of this technical vocabulary can seem confusing, his analysis is at its core quite intuitive. To put it more simply, Kant defines immorality as the ubiquitous tendency of human agents to prefer the satisfaction of their sensuous desires over the moral law. It cannot be the case that agents are morally responsible simply for having sensuous desires, for having desires is a matter of fact rather than something freely chosen. However, prioritizing the satisfaction of our desires over the moral law is free and, hence, imputable to us. The commonplace tendency to subordinate the moral law to our sensuous desires is what Kant labels the propensity to evil.

> By propensity (*propensio*) I understand the subjective ground of the possibility of an inclination (habitual desire, *concupiscentia*), insofar as this possibility is contingent for humanity in general. It is distinguished from a predisposition in that a propensity can indeed be innate yet *may* be represented as not being as such: it can rather be thought of (if it is good) as *acquired*, or (if evil) as *brought* by the human being *upon* himself. – Here, however, we are only talking of a propensity to genuine evil, i.e. moral evil, which, since it is only possible as the determination of a free power of choice … must reside in the subjective ground of the possibility of the deviation of the maxims from the moral law … [T]he will's capacity or incapacity arising from this natural propensity to adopt or not adopt the moral law in its maxims can be called *the good or evil heart*. (*Rel*, 6: 29)

Kant identifies three different stages of the propensity to evil. All three stages of the propensity can be understood generally as manifestations of a preference for sensuous incentives of inclination over the rationally stronger incentive of morality (see Wood 2014: 50). More specifically, representing increasing degrees of wickedness or grades of corruption

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

of the will, he names these three propensities: frailty (*Gebrechlichkeit*), impurity (*Unlauterkeit*) and depravity or perversity (*Bösartigkeit*, *Verderbtheit*) (*Rel*, 6: 29). It is crucial to Kant's account, and also to later criticisms of the theory, that these propensities are not degrees of evil found in actions. Rather, they are degrees of the human will's temptation to evil.

The mildest degree of evil is the frailty or weakness of the will to adhere to the moral law. Frailty is characterized by the objective recognition of the moral law accompanied by the subjective inability to adopt it as one's maxim.[11] The frail or akratic agent arrives cognitively at the proper moral conclusion, but is volitionally too weak to act upon it. As Kant describes it:

> I incorporate the good (the law) into the maxim of my power of choice; but this good, which is an irresistible incentive objectively or ideally, is subjectively the weaker (in comparison with inclination) whenever the maxim is to be followed. (*Rel*, 6: 29)

In cases of frailty, rationally inferior, i.e. sensuous or non-moral, incentives take precedence over those of reason or the moral law. The will is so *frail* that it is incapable of consistent action and often gives in to sensuous incentives.[12]

The second degree of evil is impurity. It is characteristic of impurity that 'actions conforming to duty are not done purely from duty' (*Rel*, 6: 29; see also Wood 2014: 51). In such cases, the individual acts according to the moral law, but not solely out of respect for the moral law as a duty.[13] Rather, the agent mixes moral and non-moral (i.e. either amoral or immoral) incentives and requires non-moral inclinations to subjectively reinforce morally good reasons. A person with an impure will performs actions that appear morally good, but does so only partly for the right reasons. For Kant, this kind of wrongdoing is morally worse than akrasia even though the outward empirical actions may be morally legal. For such empirical actions deceptively represent a dutiful maxim when they are in fact grounded in incentives of self-love. According to Kant, it is precisely this sort of deception that makes moral self-knowledge so difficult since the individual may think that her actions emanate from the moral law when in fact they are due to sensuous inclinations. For this reason, impure agents may not be aware of their impurity. Furthermore, it makes becoming morally good immensely more challenging because one is more likely to interpret actions that are bad as morally good, thus

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

weakening one's resolve to remain true to the demands of morality (see Frierson 2003: 112).

The third degree of evil is depravity or perversion.[14] It is 'the propensity of the power of choice to maxims that subordinate the incentives of the moral law to others (not moral ones)' (*Rel*, 6: 30). As with the other stages of evil, the non-moral incentives that are central to this propensity are incentives of self-love. However, as noted previously, evil does not arise from the mere existence of incentives of self-love in human psychology. Such incentives are a natural part of human nature, and we cannot be morally responsible for any condition or act we do not freely choose. Hence, to locate evil in the mere existence of sensuous or non-moral incentives would deprive moral responsibility for evil of its essential freedom-component; individuals must freely adopt their maxims for them to be held responsible. Therefore, Kant identifies depravity as the freely chosen *subordination* of the moral law to self-love. The human agent is literally morally depraved insofar as she freely chooses to elevate a non-moral principle – the principle of self-love – over the supreme principle of morality, thereby giving subjective priority to the former over the latter.

> [T]he difference, whether the human being is good or evil, must not lie in the difference between the incentives that he incorporates into his maxim (not in the material of the maxim) but in their *subordination* (in the form of the maxim); which of the two he makes the condition of the other. It follows that the human being (even the best) is evil only because he reverses the moral order of his incentives in incorporating them into his maxims … [H]e makes the incentives of self-love and their inclinations the condition of compliance with the moral law – whereas it is the latter that, as the supreme condition of the satisfaction of the former, should have been incorporated into the universal maxim of the power of choice as the sole incentive. (*Rel*, 6: 36)

And later in the *Religion* he states: '[S]elf-love … when adopted as the principle of all our maxims, is precisely the source of all evil (*Rel*, 6: 45).

In such cases of depravity or perversity, the individual is fully aware that incentives of morality are rationally superior to non-moral incentives. Nevertheless, the individual gives preference to the latter. Although we do not choose our inclinations or desires (and therefore cannot be morally responsible for their mere existence), we do choose what

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

significance to confer on them (see Wood 2014: 34; Card 2002: 77). In acting in a depraved manner, we reverse the ethical order of our incentives and accord our non-moral inclinations more importance than duty to the moral law. In contrast with both the frail agent who attempts to follow the moral law but is too weak-willed to do so, and the impure agent who requires sensuous motivations to act in accordance with duty, the depraved individual does not simply fail to prioritize reasons correctly (i.e. rationally and morally). Rather, the depraved person decides as a matter of principle to subordinate moral reasons to non-moral ones.[15] As Pablo Muchnik describes it: 'The value of morality is deliberately devalued, and reasons lacking (objective) justificatory power override the claim moral considerations exert on a heterogeneous will like ours' (Muchnik 2010: 161). Although Kant expresses puzzlement about *why* anyone would prioritize self-love over the moral law, it remains a fact that the tendency to do so is a fact of human nature.[16]

3. Criticisms and Replies: Can Kant's Theory of Evil Be Saved?

Kant's theory has garnered an abundance of critical attention. The primary focus of this consideration has been internal to Kant's systematic philosophy; that is, philosophers have sought to analyse and understand the coherence of the discussion of radical evil relative to his other works in practical philosophy.[17] Philosophers have been especially concerned with understanding whether Kant's insistence upon evil's radical pervasiveness, or even its innateness, in human nature leaves sufficient room for the freedom that he maintains is necessary for moral responsibility. Furthermore, there has been debate in the secondary literature regarding Kant's supposed denial of the possibility of diabolical evil, or doing evil for evil's sake.[18] Additionally, there has been some controversy regarding Kant's claim that 'we can spare ourselves the formal proof that there must be such a corrupt propensity rooted in the human being, in view of the multitude of woeful examples that the experience of human deeds parades before us' (*Rel*, 6: 32). This statement seems at odds with other statements in the *Religion* that it must be possible to construct a formal argument establishing the possibility of an evil maxim from the preconditions of such a maxim, as well as fully incompatible with his critical philosophy more generally.[19]

These concerns about the internal coherence of Kant's views are crucial to understanding his philosophical system. However, these issues are not my present focus. In addition to the issues of the 'internal' variety described in the preceding paragraph, there are a multitude of 'external' criticisms levelled at Kant's theory. These criticisms focus on the question

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

of whether Kant's theory provides a viable conception of moral evil regardless of any potential inconsistencies internal to his philosophical system. This approach is important insofar as it seeks to discover what Kant can teach us about the phenomenon of moral evil as a conspicuous feature of human interaction. The value of this kind of analysis is supported by an analogy with attention paid to Kant's categorical imperative. Although it is essential to delve into the ways in which his arguments concerning the moral law are consistent with one another, it is likewise indispensable to ask whether his discussions of the categorical imperative inform us about the nature of moral obligation and moral assessment more generally. Similarly, this essay proceeds by examining the leading 'external' criticisms levelled at Kant's theory of radical evil and evaluating the force of each criticism, in order to answer the question whether his theory can provide insight regarding the nature of moral evil and highlight important aspects concerning the moral psychology of perpetrators of evil. In order to discern whether these criticisms render Kant's theory of evil obsolete, we must address them each in turn.

Criticism One

It is a central concern of much of the recent philosophical literature concerning moral evil to distinguish evil from mundane wrongdoing.[20] This distinction is crucial to any proposed definition of evil because it demonstrates that evil is its own moral category identifying a feature of moral interaction that is irreducible to other moral categories such as bad, wrong, unvirtuous or unjust.[21]

> There are some actions, some events, that we feel the need to describe in terms of their being evil in this more exclusive sense, in which we contrast their moral status with that of everyday wrongful actions. The terrible massacres of the twentieth century, the hideous and endless ingenuity of its tortures, seem to require description in terms of evil in this exclusive sense because other kinds of moral condemnation do not capture their nightmarish horror. (Garrard 2002: 321)

The motivation behind drawing this conceptual distinction rests on the insight that some acts or persons ought to be described by a term of moral opprobrium not satisfied by terms like 'wrong' or 'bad'. Certain acts such as genocide, torture and forced starvation, and perhaps individuals capable of such acts, are not properly described as 'wrong' or 'bad' no matter how many times we preface these terms with 'very' or 'really'. The moral category of evil performs an explanatory role insofar as it properly

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

identifies the morally worst kind of acts.[22] Hence, the term 'evil' purports to have ontological and normative priority because it describes an extant category of agential action and affords apt normative judgements of acts falling under that category.

If it is indeed appropriate to distinguish between evil and mundane wrongdoing, then Kant's theory of evil, it is charged, is apparently ill equipped to do so. Under the categories of evil described by the three propensities, Kant includes acts that are, in fact, not evil outside of a Kantian context.[23] For example, recall Kant's description of the propensity for impurity. The impure individual conforms her acts to the moral law, but does so also out of sensuous incentives rather than solely out of duty to the moral law. Consequently, according to Kant, an individual is evil if she tells the truth in order to gain a good reputation rather than solely out of duty to the moral law. Yet, outside of this Kantian context, telling the truth for non-moral reasons is not morally worse than other acts of mundane wrongdoing, and therefore cannot be an example of evil. Telling the truth to establish a good reputation is altogether ordinary and not of 'nightmarish horror'. Accordingly, many of the acts that Kant refers to as evil can easily be captured by the category of mundane or ordinary wrongdoing. In fact, it is difficult to comprehend how any of the acts categorized as 'impure' by Kant could be acts that are evil in contrast to merely wrong. Since acts of impurity will always conform to the moral law (but also be done for wrong reasons), they seem to be excluded by definition from the horrendous or extremely harmful acts that characterize evil.[24] As we have already noted, for the concept of evil to have explanatory value it ought to identify acts, motives, intentions or traits that are morally worse than merely wrong ones. Otherwise, the term 'evil' is simply emotional hyperbole.[25]

Kant's conflation of evil with mundane wrongdoing is confirmed not only by the relevant passages in the *Religion* but also in the *Critique of Practical Reason*.[26] Here Kant distinguishes between two senses of 'good' and their respective negative analogues. He contrasts (a) the good (*das Gute*) and evil (*das Böse*), and (b) the good (*das Wohl*) and the bad (*das Übel or das Weh*) (*CPrR*, 5: 59–60). Under this description, to refer to something as bad in the sense of (b) is to call something harmful to one's well-being.[27] It is meant to refer to disagreeable, yet entirely accidental untoward and unfortunate circumstances. In contrast, to refer to something as evil in the sense of (a) is to call an action *morally* wrong, 'resulting not from natural contingency but from the direct action of

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

an agent's will – when she wrongly subordinates the principle of morality to the principle of self-love' (Garcia 2002: 195).

With this distinction juxtaposed with his discussion of evil in the *Religion*, we see that what Kant refers to as 'evil' should really be categorized under the broader class of 'all immoral acts'. Thus it seems that, under a strict Kantian framework, there is no unique conception of 'evil' in contrast to mere or mundane immoral acts.[28]

Reply One

Is Kant guilty of conflating mundane wrongdoing and evil? Briefly put: yes. As just discussed, Kant's theory of evil is in fact a discussion of all of the possible ways of committing immoral acts generally speaking. One commits immoral acts through the propensities to frailty, impurity or depravity. Although it is possible to identify acts of evil as one kind of immoral act falling under these broad categories, Kant does not carve out a special category of evil in distinction to mundane wrongdoing.

Nevertheless, it would be uncharitable to Kant to describe this conflation as culpable. We need to keep in mind Kant's main goal in the *Religion*. His principal objective is to explain religion in terms of rational morality. As indicated by the title, the book purports to explore what reason can provide us in the way of religion without resorting to divine revelation. In the Preface to the second edition Kant writes:

> Regarding the title of this work (since doubts have been expressed also regarding the intention hidden behind it) I note: Since after all, *revelation* can at least comprise also the pure *religion of reason*, whereas conversely, the latter cannot do the same for what is historical in revelation, I shall be able to consider the first as a *wider* sphere of faith that includes the other, a *narrower* one, within itself … the philosopher, as purely a teacher of reason (from mere principles *a priori*), must keep within the inner circle … (*Rel*, 6: 12)

Kant's purpose is precisely indicated by the book's title. He intends to investigate religion within the bounds of reason. Since religious belief supported by revelation is empirical in its source, it is the appropriate content of theological, but not philosophical inquiry. As a philosopher, Kant's goal is to remove 'everything empirical' (*Rel*, 6: 12–13) from his inquiry into religion and religious doctrine. (Nevertheless, he believes

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

that between his methodology and biblical scripture 'there is not only compatibility but also unity, so that whoever follows the one (under the guidance of moral concepts) will not fail to come across the other as well', *Rel*, 6: 13.)

What is more, his goal is to pursue this investigation for a specific audience of late eighteenth-century German Lutherans. Kant expects his audience to hold certain religious beliefs that merge with their life as rational moral agents. His aim is to convince them that there is 'no fundamental conflict between these two aspects of their moral life' (Wood 2014: 31). To this end, he adopts certain language that was more common in a religious rather than a secular philosophical setting. It is understandable that he chose to focus on evil (*das Böse*), which plays a central role in Christianity generally and Lutheranism in particular. Finally, Kant's discussion is less focused on what the nature of evil is, but on transcendental evil – how evil occurs, how its occurrence is compatible with freedom and how we can understand it through reason rather than revelation.[29]

With these facts in mind, we see that it was simply not Kant's goal to explicate the difference between evil and mundane wrongdoing. It is perhaps anachronistic in a way to expect him to have pursued a conceptual distinction that has proven to be philosophically significant through contemporary discussions of moral evil. Rather than asking *if* Kant distinguished between evil and mere wrongdoing, it is more philosophically generous to ask whether Kant *could* have distinguished between these two concepts given what he says about evil specifically and immorality more generally.

It is possible to answer this question affirmatively. Although Kant left the conceptually possible distinction between evil and mundane wrongdoing unexplored, Ernesto Garcia argues that one can reasonably modify the Kantian ethical theory in order to locate 'a fundamental *qualitative* difference between evil and more ordinary immoral actions … in which … a "material" (as opposed to purely "formal") difference exists between the respective maxims of the immoral and the evil agent' (Garcia 2002: 194). It is a hallmark of Kantian ethics that the form of the maxim and not its matter determines the moral worth of an action (*G*, 4: 399–400). The moral law obligates us formally to recognize humanity as an objective end comprised of all rational beings and regarded as an 'end in itself' (*G*, 4: 429). Conversely, immoral actions have the character of positing some 'material end' that is not determined by pure reason, but rather *a posteriori* by natural incentives (*G*, 4: 427–8). The principle that ought to govern our action, namely, the principle that recognizes the objective end of humanity as an

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

end in itself (i.e. the categorical imperative), is subordinated to the principle of self-love. As a result, the agent uses the humanity of others as a mere means to the realization of some concrete state of affairs.

According to Garcia, evil actions distinguish themselves from merely immoral ones by emanating from a specific kind of material end. Immoral actions are characterized by using the humanity of others merely as a means to some contingent end, but 'in evil actions we seek out to directly violate the humanity of another person itself' (Garcia 2002: 201). Whereas the moral law commands us to regard humanity as the necessary formal end of moral action, in evil action 'the object of our willing *just consists* in the mistreatment of the humanity of another person *qua* human'; evil action has as its material end 'the violation of humanity itself' (ibid.). Understood in this way, Kant's practical philosophy can be interpreted to make room for a distinction between evil and mere wrongdoing.

When we consider acts like genocide, torture and forced starvation, the direct violation of humanity itself is an apt characterization and pinpoints why such horrendous acts deserve to be categorized as evil rather than simply wrong. If the direct violation of humanity is central to what makes an act evil, then Kant's claim that recognizing humanity as an objective end in itself is central to morality (and immorality) indicates that his theory is indeed an appropriate one for understanding the nature of evil. A unique kind of direct violation of that which is central to morality distinguishes evil from mundane wrongdoing. Thus it seems as if Kant can provide us with the conceptual tools to locate the identifying characteristic of evil as distinct from mere wrongdoing. If so, then we can conclude that Criticism One is not substantial enough to persuade us to abandon Kant's theory of evil. Be that as it may, Reply One is admittedly not fully satisfying. Although it provides us with the best way to handle the criticism, we are nevertheless left with the impression that we have merely evaded the issue rather than solved it directly. As a result, Reply One is not sufficient on its own to convince Kant's critics of the continued relevance of his account. For this, we need to move on to the next two criticisms and my responses.

Criticism Two

Many critics argue that Kant's theory of evil is overly simplistic and narrow in its reduction of evil to the subordination of the moral law to self-love. With Kant's account in mind John Kekes writes:

> Most of the explanations given in the framework of the religious or the Enlightenment world view assume that evil has a single

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

> cause. Evil, however, has many causes: various human propensities; outside influence on their development; and a multiplicity of circumstances in which we live and to which we must respond. Because these causes vary with person, time, and place, an attempt to find *the* cause of evil is doomed. There is no explanation that fits all or even most cases of evil. (Kekes 2005: 4)

Even if we grant to Kant that the prioritization of self-love is a probable root of evil, it seems implausibly reductionist to claim that it is the only one. The seemingly unending list of cruelties pervading human history are too numerous and varied to attribute to a single cause. A quick glance at this history indicates that evil can be done out of malice and hate, but also greed, boredom, ambition, revenge, fear and obedience.[30] Therefore, to claim that evil is attributable only to the subordination of the moral law to self-love is to ignore the manifold ways that evil occurs in moral interaction.

Reply Two

There are two responses to this criticism that may help save Kant's theory. First, we must precisely identify Kant's understanding of the root of evil – that is, the subordination of the moral law to the principle of self-love.[31] *Selbstliebe* is most commonly translated into English as self-love, and on occasion as self-interest.[32] Rather than investigate the propriety of the translation, I find it more revealing to examine its use in order to discern its meaning. 'Self-love' understood in a limited, non-Kantian, and perhaps more common, way refers mainly to a self-indulgent hedonist or egoist. For Kant, in contrast, it refers to the whole array of non-moral, natural incentives that navigate towards happiness.[33] To see this, recall that Kant's definition of evil unfolds within the framework of his analysis of the conflict or struggle between two very general categories of incentive: the moral and the non-moral. 'Inclination' is simply the Kantian term for non-moral incentives, and 'self-love' is a general term that is meant to encompass the pursuit of objects of inclination. As Allen Wood writes, these terms are 'merely placeholders for whatever non-moral incentives might be chosen in preference to those of morality. Kant is not imposing any limits on what one can have an inclination to will' (Wood 2014: 36; see also Louden 2000: 138–9). Since Kant uses 'self-love' as a broad term covering *all* non-moral incentives, and argues that evil comes about due to prioritizing non-moral incentives over the moral law, his theory can apparently accommodate the notion that there are many roots of evil.

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

Although this first response to Criticism Two is convincing on several levels, it remains vulnerable to an objection. One might resist the interpretation offered in the preceding paragraph and insist that all sensuous inclinations can indeed be reduced to the single root of self-love. It is not psychologically far-fetched to claim that the common motivations behind acts of evil – e.g. greed, boredom, ambition, revenge, fear and obedience – are fundamentally manifestations of self-love. Addressing this possibility, I offer an analysis that reinterprets Kant's theory using the terminology of contemporary moral philosophy and suggest in what way self-love alone is a plausible origin of evil.

The following argument proceeds in two main steps. First, I suggest that evil may be most adequately understood as resulting from an agent's *rendering irrelevant overriding normative reasons for action*. Second, I argue that insofar as such overriding reasons are second-personal reasons, self-love explains why an agent may render them irrelevant.

We begin this discussion by assenting to the quite credible suggestion that an agent's reasons for acting are germane to actions of evil. Eve Garrard agrees and argues that evil ought to be understood in terms of the phenomenon of 'silencing' as it relates in two distinct ways to reasons.[34] 'Metaphysical silencing' occurs when in the presence of reasons that make an action right (or good) other considerations cease to be reasons at all. Alternatively, 'psychological silencing' is 'a total failure [on the part of an agent] to see that certain considerations are reasons at all' (Garrard 2002: 329–30).

> The evil act is done by one for whom (at least some of) the considerations that tell against his committing this wrongful act, are silenced altogether. The sufferings of his victims, along with other considerations such as their rights, play no part in his practical deliberations. They count for nothing at all. And it is this silencing, this inability to hear the victims' screams as significant, that accounts for the peculiar horror that we feel when we contemplate these evil acts and their agents. (Garrard 1998: 53–4)

As Garrard describes it, an evil action occurs when an agent psychologically silences significant reasons against her acting. Reasons are significant when they serve to metaphysically silence other considerations that in differing circumstances could have reason-giving force. In other words, an agent performs an evil act when she psychologically silences reasons which are themselves metaphysical silencers.

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

The term 'metaphysical silencer' seems conceptually (and metaphysically) loaded, but the rationale behind it is simple: certain considerations lose their reason-giving force in the presence of strong reasons to do the morally right thing. For example, the possibility of saving money is an acceptable reason for buying one box of cereal over another.[35] In contrast, it is not a reason at all to refrain from buying a much-needed medicine for my child who might die a painful and prolonged death without the medicine. Some reasons are so powerful that they have the effect of metaphysically silencing alternative considerations, which in other contexts might have reason-giving force. In the context of saving my child's life, saving a modest amount of money is simply no reason at all.[36]

I find Garrard's theory to be compelling, but her talk of 'silencing' is unnecessarily confusing, psychologically and metaphysically onerous, and difficult to prove. Rather than utilize the notion of 'psychological silencing', or refer to the capacity of reasons to 'metaphysically silence' other considerations, I believe we can capture the same insights while avoiding the excessive implications by using the phrases 'rendering irrelevant' and 'overriding'.[37] On this account evil occurs when an agent renders irrelevant, that is, consciously or perhaps even unconsciously rejects, the presence of overriding normative reasons against her acting. Reasons are overriding when they turn alternative potential reasons into mere considerations, depriving them of any reason-giving force. To put the proposed idea more simply, an agent performs an evil act when she renders irrelevant overriding normative reasons for action.

Two questions arise. What kinds of reasons are normatively overriding? And how could an agent render such reasons irrelevant? To answer the first question we first observe that moral agents find themselves in relationships of moral address and demand with one another.[38] Moral agents address each other normatively by making claims, requests or demands on others and responding to others' claims, requests or demands. Normative claims, requests or demands from others provide a moral agent with agent-relative reasons for action. Stephen Darwall calls such reasons second-personal reasons:

> Second-personal reasons are invariably tied to a distinctively second-personal kind of *practical authority*: the authority to make a demand or claim. Making a claim or demand as valid always presupposes the authority to make it and that the duly authorized claim creates a distinctive reason for compliance (a second-personal reason). (Darwall 2006: 11)

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

Agents in the moral community are moral agents precisely insofar as they are addressees and addressors of moral claims, requests and demands. In other words, moral agents engage in reciprocal relations of moral address, and the claims, requests and demands that make up this address provide agents with reasons for action.

In contexts of great moral import such reasons are overriding. For example, under non-moral circumstances there are reasons for me to perform actions that satisfy my individual preferences or bring about some degree of pleasure for me. However, if the means to achieve these ends require the torture of another person, then clearly the second person's demands not to be tortured override my reasons for performing an action that brings about some degree of pleasure for me. Indeed, because second-personal reasons are by definition normatively significant they will often override non-moral reasons, and they always override in circumstances of great moral import.[39]

Despite the fact that second-personal reasons are often overriding, some individuals render these reasons irrelevant. I propose that we can explain this process through an appeal to self-love. Individuals who have an immutable preference for the satisfaction of their own inclinations are individuals who are prone to reject second-personal reasons to act. The demands of others play absolutely no role in their practical reasoning. These individuals are moral solipsists who prioritize first-person considerations even when such considerations have lost all reason-giving force.[40] They allow considerations of self-love to render second-personal reasons irrelevant. Hence, evil acts are performed when an agent, out of self-love, renders irrelevant overriding second-personal reasons for action.

Understood in this way, self-love is indeed the root of moral evil because it allows the normative claims, requests and demands of others to go unaddressed. Since the addressing of these claims, requests and demands is the essence of moral agency, and the meeting of these claims, requests and demands when appropriate is the core of morally praiseworthy action, we can conclude that Kant is correct to claim that the prioritization of self-love over the moral law is the single cause of immoral and evil action.

Criticism Three

It has been widely objected that by focusing solely on the intelligible adoption of a supreme principle for action, Kant's theory fails to give heed to the most salient feature of an evil act: the actual empirical harm and suffering of the victim. By defining evil solely as weakness of will,

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

impurity of will and depravity of will – the subordination of the moral law to the principle of self-love – an agent's action would not need to cause any suffering or produce any victims at all for the act to be evil. Can evil be defined without the infliction of suffering? Daniel Haybron suggests that someone might qualify as evil despite never bringing about any harm:

> Take the vilest person you can imagine and make her a quadriplegic with no ability to communicate: living in silent spite, she wishes nothing more than the greatest suffering for her fellow creatures, and takes the greatest joy in witnessing others in agony. The handicap makes her less dangerous, but it scarcely makes her a better person. Nor does it make her anything better than evil. (Haybron 2002: 264)

Certainly this individual is despicable. Yet, contrary to his intention, Haybron's terminology betrays a significant characteristic of evil. As he describes her, the individual is not evil, but vile. Not only her misanthropy, but the joy she feels in witnessing others' suffering makes her contemptible. However, the inability to cause and inflict suffering on others frees her from the moniker of evil. As Paul Formosa writes: 'there is an important moral difference between merely wishing and fantasizing about evil … and actually intending and inflicting evil. While the quadriplegic remains a vile human being, she is not an evil person until she turns her evil fantasies into evil intentions and evil acts' (Formosa 2007: 235).

Indeed, numerous philosophers argue that harm, specifically a certain kind or degree of excessive or life-wrecking harm, is a necessary condition of an evil act.[41] For example, Peter French notes that 'victimization is the identifying characteristic of evil' (French 2011: 62–3). Arne Johan Vetlesen argues that evil is to 'intentionally inflict pain and suffering on another human being, against her will, and causing serious and foreseeable harm to her' (Vetlesen 2005: 2). Each of these theories identifies a necessary harm component in evil acts that characterizes the act as evil rather than merely wrong. Even if the motives, intentions and maxims of an agent are relevant to discussions of the nature of an evil act, the harm inflicted on the victim is as well. To see this, just imagine the silly example of someone stealing a candy bar from a store out of sheer malice. It is clear that the malicious intention is not sufficient to characterize the act as evil. The element of extreme, prolonged or life-wrecking harm is missing.[42] If empirical harm or suffering is a necessary element of evil,

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

and Kant excludes this element from his theory, then his approach is incomplete at best, and false at worst. Criticism Three makes *prima facie* the strongest case in favour of rejecting Kant's account of evil. A theory of evil without victims is no theory of evil; it is a theory concerning the potential for evil.

Reply Three

There are two responses that defenders of Kant's theory can develop. The first response requires a brief review of Kant's notion of a maxim to indicate that he can or does accommodate harm and the suffering of evil's victims into his account of evil. The second response explains why a focus on the quality of will of the evildoer is essential to an accurate theory of evil.

The suggestion that Kant's emphasis on the intelligible adoption of maxims reflects a complete disregard for the significance of empirical actions, and *ipso facto* empirical harm, is a caricature of his view and is simply false. Maxims, according to Kant, are subjective volitional principles of action (*G*, 4: 399–400; cf. Wood 1999: 40). In other words, they are principles that motivate the will and are practical by their very nature.

The confusion arises by conflating Kant's insistence that the moral worth of an action is to be determined solely by the formal maxim that gives rise to it, and not in the end to be attained by it, with the idea that a noumenal maxim stands somehow in isolation from its phenomenal counterpart – the action (see *G*, 4: 399–401). However, the mere fact that the moral worth of an action is determined by its formal maxim does not divorce a maxim from its subsequent action. Insofar as a maxim 'is a general rule by which [an individual] will conduct himself' (*Rel*, 6: 24), it is inherently connected to an individual's conduct or action. Even if the moral worth of a maxim is determined purely by virtue of its form, 'every maxim involves an end which serves as its matter' (Wood 1999: 52; see *G*, 4: 436, *CPrR*, 5: 21).

Consequently, even if it is true that in his theory of evil Kant's main concern is with the quality and corruption of the will of the agent, he is not blind to the simple fact that empirical actions follow directly from an agent's maxims.[43] Because empirical actions follow from the adoption of maxims, it is almost tautological that the adoption of maxims is significantly phenomenal. As a result, it is unfairly dismissive of Kant's view to argue that his theory holds no regard for the actual harm suffered by victims of evil.

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

One might reply that even if Kant's theory does pay heed in this way to empirical suffering, there is no doubt that his analysis remains principally transcendental rather than empirical – that is, concerning those characteristics that make the doing of evil possible. By identifying the transcendental features of evil, he directs our attention to evil's possibility rather than its actual occurrence. Nevertheless, I reply that a focus on the former does not entail believing that the latter is inconsequential. In fact, Kant attends to empirical evil directly in his discussion of the 'vices of culture'. He observes that we have constant anxiety that other people will attain superiority over ourselves. Through a continuous comparison with others there arises a desire for superiority over them. Out of this rivalry emerge 'envy, ingratitude, joy in others' misfortunes, etc.' (*Rel*, 6: 27), all of which occur in the phenomenal world.[44]

Even so, there is a significant reason why Kant does not focus his account of evil on empirical harm. We must keep in mind not simply *that* Kant's account is transcendental rather than empirical, but also on the reasons *why*. What is so morally troublesome about the vices of culture, and all of evil, is that they hide behind reason and are perfectly compatible with good empirical conduct.[45] Recall Kant's emphasis on the ways in which immorality is cloaked in self-deception. The agent may perform morally legal acts while being motivated by immoral incentives. Far from trivializing the actual suffering of victims of evil, Kant seeks to end it by identifying its source and possibility. The prevention of empirical suffering is only the beginning of the struggle against evil, not its conclusion. For this reason, his approach to evil focuses on individuals' quality of will and evil's transcendental conditions.

This judgement renders Kant's theory highly relevant for contemporary conceptions of evil and morality, and identifies a crucial point relevant to discussions of evil and the moral psychology of perpetrators of evil. Morally legal acts – that is, acts that outwardly adhere to the moral law – are not constitutive of morality because they do not reflect the moral quality of an agent's character. The absence of empirical suffering is not an indication of moral goodness because it is merely contingent whether an individual acts according to the moral law when she has not prioritized the moral law in her will. This leads to the possibility, if not the likelihood, that the individual will act contrary to morality in different circumstances. Consequently, appearances of morality are no guarantee against future occurrences of evil. Especially insofar as an individual can be deceived about the moral goodness of her own character, evildoing gains a tenacious foothold when an agent fails to act out of duty to the moral law and acts only in accordance with it.

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

This insight acquires both gravity and credibility when we consider it alongside the large-scale mass atrocities of the past century. Empirical studies have demonstrated that the people who perpetrated the Holocaust or the genocide in Rwanda were ordinary individuals rather than diabolical monsters.[46] Nevertheless, it remains puzzling how ordinary people can so easily commit extraordinary acts of evil. Although Kant and his contemporaries could not have fathomed the massacres that in large part defined the twentieth century, his theory of evil helps us understand the ease with which ordinary individuals became perpetrators of the morally worst kinds of acts. Without a morally good moral character, the fact that an agent has successfully acted in accordance with the moral law in the past is no reason to think she will do so in the future. Rather, it is good reason to think that she will not.[47] Precisely because people are not diabolical, but rather tend to be mistaken about the moral quality of their wills, are easily motivated by sensuous incentives and commonly misunderstand the nature of morality and evil, they can become perpetrators of evil.

Contemporary criticisms of the Kant's theory often miss this point and its relevance for the prevention of future evil, thereby neglecting the importance his theory still holds for understanding the phenomenon of evil. The constant struggle in the human will regarding the prioritization of moral, non-moral or immoral incentives is definitive of moral character. Whichever incentives win out in this struggle constitute one's moral character. Without a moral character dedicated to morality, humans will tend to act in evil ways.

4. Conclusion

I have endeavoured to save Kant's theory of evil from outright rejection. To this end, I addressed three prevalent criticisms of his theory found in the contemporary philosophical literature concerning moral evil. First, I demonstrated that, although Kant did not intend to distinguish evil from mere wrongdoing, he has the conceptual tools available to do so. One can modify Kant's ethical theory and claim that an evil maxim is distinct from a merely wrong one in that the former sets as its material end the destruction of the humanity of others. Second, I noted that Kant's use of the term 'self-love' is a placeholder for all non-moral incentives. Further, the notion that self-love is the single fundamental cause of evil ought not be so readily dismissed. We can plausibly define an evil act as one performed by an agent who through self-love renders irrelevant overriding normative reasons. Finally, the suggestion that Kant shows no regard for empirical actions is simply false. The adoption of maxims, even though

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

noumenal, is inherently connected to an agent's phenomenal conduct. Furthermore, we gain insight into the volitional struggle among moral, non-moral and immoral incentives that makes evil possible as well as understand that outward appearances of moral goodness without a morally good character do not safeguard against evil's future occurrence. Although Kant's theory of evil is not without its philosophical difficulties, we can nevertheless indicate its relevance for contemporary discussions of moral evil. Rather than being dismissed as lacking sufficient depth, Kant's position regarding the subject of radical evil deserves to be reconsidered alongside contemporary theories that attempt to understand evil's nature and pervasiveness.[48]

Notes

1 For notable exceptions, see Wood (2014), Formosa (2007), Louden (2000: 132-9).

2 Card (2002: 73; 2010a: 43), Singer (2004: 187, 209), Russell (2014: 20, n. 5). The following texts attempt to make explicit a distinction between mere wrongdoing and evil: Thomas (1993), Garrard (2002), Perrett (2002), de Wijze (2002), Morton (2004), Kekes (2005), Vetlesen (2005), Formosa (2008), Barry (2012), Calder (2013).

3 Silber (1991), Bernstein (2001: 84; 2002: 42), Kekes (2005), Formosa (2008), Card (2010b).

4 Card (2002, ch. 4; 2010b: 74). Additionally, the following texts implicitly reject Kant's theory by asserting that evil is defined by virtue of the harm it causes rather than the psychology of the perpetrator: Kekes (1990: 4; 2005: 1), Vetlesen (2005), French (2011: 61, 95), Russell (2012, 2014). See also Gowans (1994: 117–54). He argues that Kant's philosophy in general is preoccupied with an agent's moral purity and not concerned with other individuals with whom the agent interacts.

5 I will use the following abbreviations and translations. *CPrR* = *Critique of Practical Reason*, *G* = *Groundwork for the Metaphysics of Morals* (Kant 1996); *Rel* = *Religion within the Boundaries of Mere Reason* (Kant 1998). Pagination is by volume/page from the Akademie edition of Kant's works.

6 Kant calls the predisposition to respect the moral law that of 'personality'. We have two other predispositions as well: to animality and to humanity. The predisposition to animality provides for our preservation of the species, and the predisposition to humanity produces the inclination to acquire worth in the eyes of others. Each of the three predispositions reflects a fundamental aspect of human nature and is originally oriented to achieve some good end. That is, they are considered good in themselves.

7 One could easily assume that the use of the word 'radical' by Kant is meant to refer to evil's severity. However, Kant's use of the term refers to its pervasiveness rather than its magnitude. The original German reflects this: Kant refers to *das radikal Böse* rather than *das radikale Böse*. The first usage reveals adverbial rather than adjectival use. That is, Kant's use of the word 'radikal' does not modify the noun 'Böse' (evil). For more on 'radical' as pervasiveness rather than intensity, see also Bernstein (2002: 20–8), Loudon (2000: 135–6).

8 See Munzel (1999: 57–70) for an analysis of *Gesinnung* in relation to good moral character. See Palmquist (2015) for a detailed account of *Gesinnung* as volitional conviction rather than a disposition or attitude.

9 See Michalson (1990: 30–51) for a detailed discussion of the relationship between predisposition and propensity, and (1990: 52–70) for the relationship between disposition and propensity.

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

10 Kant actually gives two reasons for calling evil radical. The first reason is as described above. The second reason is connected to the first, but has a different focus. Kant insists that (evil) is 'not to be extirpated through human forces' (*Rel*, 6: 37). Nevertheless, Kant adds that: 'Yet it must be equally possible to overcome this evil, for it is found in the human being as acting freely' (*Rel*, 6: 37). Kant's discussion of overcoming evil through moral conversion, or moral regeneration as it is sometimes referred to in the literature, occupies *Religion* II, but a robust analysis of this aspect of his theory falls outside the scope of this essay.

11 I thank one of the anonymous reviewers for helpful remarks on this point.

12 For a critical examination of the Kantian notion of frailty, see Frierson (2014: 232–58).

13 Actually, this kind of legalistic morality – acting according to duty but not out of duty – is not excluded in the next stage of evil, perversity/depravity, either. I am grateful to Federica Basaglia for her comments on this point. I return at the end to the importance of this aspect of Kant's theory for contemporary theories of evil.

14 Kant refers to this degree of evil in different ways. See *Rel*, 6: 30.

15 Nevertheless, Kant insists that this kind of evil is not doing evil for evil's sake, which is demonic and, by definition, humanly impossible. Rather, the depraved individual has a perverse conception of the good.

16 For Kant, since there is quite literally *no reason* to adopt the maxim involved with evil, it is inscrutable why anyone would do so. See also Grimm 2002: note 22.

17 Some of the more recent and influential examples include: Wood (1999: 283–90), Louden (2000: 132–9), Grimm (2002), Allison (2002), Caswell (2006), Formosa (2007), Anderson-Gold and Muchnik (2010), Muchnik (2010), Kemp (2011), Silber (2012), Firestone and Jacobs (2013), McMullin (2013), Michalson (2014).

18 In addition to the sources listed in the preceding note, see also Silber (1991), Allison (1996: 169–82), Bernstein (2001, 2002), Card (2002: 84–7), Formosa (2009), Wood (2014).

19 Again, in addition to the sources listed in the previous two notes, see also Michalson (1990).

20 See Thomas (1993), Card (2002), Garrard (2002), Perrett (2002), de Wijze (2002), Morton (2004), Singer (2004), Kekes (2005), Vetlesen (2005), Formosa (2008), Barry (2012), Garrard and McNaughton (2012) Calder (2013), Russell (2007, 2012, 2014).

21 There is some disagreement about whether the difference between merely wrong and evil is qualitative or quantitative. For an overview and analysis of this debate, see Russell (2007).

22 Steiner (2002) raises the possibility of defining evil as a negative counterpart to the supererogatory. Just as the supererogatory categorizes those acts that go beyond the good, 'evil' refers to acts that go beyond the merely bad. Both supererogatory and evil acts involve intensified value or disvalue, respectively. However, looking more closely at deontic categories in relation to an axiological scale will show that explaining evil as the negative counterpart to the supererogatory fails due to the familiar problem of identifying a class of actions that has supreme value but is not morally required. Supererogatory acts are by definition not obligatory (weakly permissible), and are certainly not forbidden, and therefore they fall under the category of the strongly permissible. Here the disanalogy with evil acts arises. Evil acts are both the most disvaluable ones *and* completely forbidden. There is a symmetry between the two on an axiological scale but an asymmetry in terms of the deontic categories.

23 See also Kekes (1990: 4), Card (2002: 3). See Gressis (2009) for an attempt to defend Kant against this charge.

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

24 This aspect of this criticism, namely that acts that conform to the moral law cannot be evil, is connected to Criticism Three, which accuses Kant of ignoring empirical harm. Since a complete response to this point requires the discussion that falls under my response to Criticism Three, I address this issue in full at the end of the essay.

25 See Clendinnen (2002) and Cole (2006) who argue against the propriety of evil as a moral concept.

26 One might object that the present discussion is merely scholastic with little philosophical import. There are two responses to such a worry – one general and the other specific to the current analysis. First, scholastic discussions can have drastic philosophical import insofar as the former can be prerequisites of the latter. For how can we address the philosophical import of an argument or theory if we are not clear about what position that argument or theory is presenting or defending? Second, my goal in this essay is to address the main criticisms in the secondary literature of Kant's theory of evil. One such criticism is that Kant does not distinguish between mere wrongdoing and evil. Hence, so the criticism goes, his theory is not a theory of evil at all. In assessing the validity of this criticism, we can gain philosophical insight into the metaethical nature of moral evil both in Kantian and non-Kantian senses of the term.

27 Referred to as 'illbeing' in the Cambridge Edition.

28 Contrast my view with that of Garcia (2002).

29 I return to the significance of the distinction between empirical and transcendental evil at the end of the essay.

30 I do not intend this list to be exhaustive.

31 For a recent discussion of the distinction between self-love and self-conceit in Kant's practical works, see Moran (2014).

32 Card (2002) refers to *Selbstliebe* as self-interest as does Vetlesen (2005: 128). However, elsewhere Kant uses the term *Eigennutz* to refer to self-interest (*G*, 4: 398). Most of the secondary literature refers to *Selbstliebe* as self-love, and I will follow this tradition.

33 I thank one of the anonymous reviewers for helping to clarify this claim.

34 Borrowing a distinction first made by McDowell (1978). She also presents this argument in Garrard (1998).

35 This example is Garrard's, bearing a slight modification.

36 Garrard insists that metaphysical silencing is not equivalent to the outweighing of one reason by another more powerful or significant one. See especially (2002: 330) and (1998: 53–4). Indeed, reasons cannot be merely outweighed if they are silenced completely. However, I am not convinced by her argument in this regard. An evil act can be performed by an agent for whom significant normative reasons are unjustifiably outweighed rather than completely silenced. Even Garrard's discussions of partial silencing seem to favour this view. For a recent collection concerning the weighing and outweighing of reasons, see Lord and Maguire (2016).

37 There is a further conceptual difficulty with Garrard's account if one relies on the notion of silencing. If some reasons are metaphysical silencers due to their reason-giving force, then how is it possible for an agent to psychologically silence them? If other considerations have been metaphysically silenced, then they are not metaphysically available for assent. Therefore, the notion of psychological silencing is difficult to make sense of in relation to the notion of metaphysical silencing.

38 Strawson (1993), Darwall (2006), Shoemaker (2007, 2015), McKenna (2012).

39 I accept that there may be situations in which non-moral or prudential reasons override second-personal reasons. For example, the reason I have to work late in order to finish writing an article may override the reason I have to call the friend I have not spoken to in

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

ages. However, in contexts of great moral import, second-personal reasons will override other reasons.

40 In the context of P. F. Strawson's terminology, we can also think of the evildoer as lacking the ability to experience, or perhaps rejecting the importance of, the vicarious and self-reactive attitudes and recognizing only the personal ones. The evildoer recognizes only that she makes demands on others. This raises difficult questions of moral responsibility. See Strawson (1993).

41 See Kekes (1990: 4, 2005: 1), Card (2002: 3), Formosa (2008), Goldberg (2016).

42 Nevertheless, harm is not always sufficient to characterize an act as evil. Paul Formosa captures this insight with an illustrative example. Imagine a drunk driver whose erratic driving causes another car to swerve and crash killing the family of five inside (Formosa 2008: 221). The individual acted culpably and his actions caused severe harm, yet we would hesitate to categorize his actions as evil. Accordingly, severe or excessive harm appears to be a necessary, albeit not sufficient, element of an evil act.

43 Indeed, Kant considers appearances to be morally important because they are expressions of a free will.

44 See also Anderson-Gold (2010). She argues that Kant's theory of radical evil is conceptually equipped to make sense of empirical evil, in particular crimes against humanity.

45 I thank one of the anonymous reviewers for helpful comments on this part of the discussion.

46 See Arendt (1964), Browning (1992), Clendinnen (2002), Straus (2006).

47 I have benefited from a discussion with Axel Hutter concerning this point.

48 I thank the anonymous reviewers for their careful reading and insightful comments.

References

Allison, Henry E. (1996) *Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy*. Cambridge: Cambridge University Press.

—— (2002) 'On the Very Idea of a Propensity to Evil'. *Journal of Value Inquiry*, 36, 337–48.

Anderson-Gold, Sharon, and Pablo Muchnik (eds) (2010) *Kant's Anatomy of Evil*. Cambridge: Cambridge University Press.

Arendt, Hannah (1964) *Eichmann in Jerusalem: A Report on the Banality of Evil*. New York: Viking Press.

Barry, Peter Brian (2012) *Evil and Moral Psychology*. New York: Routledge.

Bernstein, Richard J. (2001) 'Radical Evil: Kant at War with Himself'. In María Pía Lara (ed.) *Rethinking Evil: Contemporary Perspectives* (Berkeley, CA: University of California Press), pp. 55–85.

—— (2002) *Radical Evil: A Philosophical Interrogation*. Cambridge, MA: Polity Press.

Browning, Christopher (1992) *Ordinary Men: Reserve Police Battalion 101 and the Final Solution in Poland*. New York: Harper Collins.

Calder, Todd (2013) 'Is Evil Just Very Wrong?' *Philosophical Studies*, 163, 177–96.

Card, Claudia (2002) *The Atrocity Paradigm: A Theory of Evil*. Oxford: Oxford University Press.

—— (2010a) *Confronting Evils: Terrorism, Torture, Genocide*. Cambridge: Cambridge University Press.

—— (2010b) Kant's Moral Excluded Middle'. In Sharon Andersen-Gold and Pablo Muchnik (eds), *Kant's Anatomy of Evil* (Cambridge: Cambridge University Press), pp. 74–92.

Caswell, Matthew (2006) 'Kant's Conception of the Highest Good, *Gesinnung*, and the Theory of Radical Evil'. *Kant-Studien*, 97, 184–209.

Clendinnen, Inga (2002) *Reading the Holocaust*. Cambridge: Cambridge University Press.

Cole, Phillip (2006) *The Myth of Evil*. Edinburgh: Edinburgh University Press.

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

Darwall, Stephen (2006) *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press.
de Wijze, Steven (2002) 'Defining Evil: Insights from the Problem of Dirty Hands'. *The Monist*, 85 (2), 210–38.
Firestone, Chris L., and Nathan Jacobs (2013) *In Defense of Kant's Religion*. Bloomington, IN: Indiana University Press.
Formosa, Paul (2007) 'Kant on the Radical Evil of Human Nature'. *Philosophical Forum*, 38 (3), 221–45.
—— (2008) 'A Conception of Evil'. *Journal of Value Inquiry*, 42, 217–39.
—— (2009) 'Kant on the Limits of Human Evil'. *Journal of Philosophical Research*, 34, 189–214.
French, Peter A. (2011) *War and Moral Dissonance*. Cambridge: Cambridge University Press.
Frierson, Patrick (2003) *Freedom and Anthropology in Kant's Moral Philosophy*. Cambridge: Cambridge University Press.
—— (2014) *Kant's Empirical Psychology*. Cambridge: Cambridge University Press.
Garcia, Ernesto (2002) 'A Kantian Theory of Evil'. *The Monist*, 85 (2), 194–209.
Garrard, Eve (1998) 'The Nature of Evil'. *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, 1, 43–60.
—— (2002) 'Evil as an Explanatory Concept'. *The Monist*, 85 (2), 320–36.
Garrard, Eve, and David McNaughton (2012) 'Speak No Evil?' *Midwest Studies in Philosophy*, 36, 1–17.
Goldberg, Zachary J. (2016) 'Evil, "Evil", and Taking Responsibility'. In Birgit Recki (ed.), *Wozu ist das Böse gut?* (Münster: Mentis), pp. 17–36.
Gowans, Christopher W. (1994) *Innocence Lost: An Examination of Inescapable Wrongdoing*. Oxford: Oxford University Press.
Gressis, Robert (2009) 'How to Be Evil: The Moral Psychology of Immorality'. In Pablo Muchnik (ed.), *Rethinking Kant*, vol. 1 (Newcastle: Cambridge Scholars), pp. 191–216.
Grimm, Stephen R. (2002) 'Kant's Argument for Radical Evil'. *European Journal of Philosophy*, 10 (2), 160–77.
Haybron, Daniel (2002) 'Moral Monsters and Saints'. *The Monist*, 85 (2), 260–84.
Kant, Immanuel (1996) *Practical Philosophy*. Trans. and ed. Mary J. Gregor. Cambridge: Cambridge University Press.
—— (1998) *Religion within the Boundaries of Mere Reason and Other Writings*. Trans. and ed. Allen W. Wood and George Di Giovanni. Cambridge: Cambridge University Press.
Kekes, John (1990) *Facing Evil*. Princeton: Princeton University Press.
—— (2005) *The Roots of Evil*. Ithaca, NY: Cornell University Press.
Kemp, Ryan (2011) 'The Contingency of Evil: Rethinking the Problem of Universal Evil in Kant's *Religion*'. In Pablo Muchnik (ed.), *Rethinking Kant*, vol. 3 (Newcastle: Cambridge Scholars), pp. 100–23.
Lord, Errol, and Barry Maguire (eds) (2016) *Weighing Reasons*. Oxford: Oxford University Press.
Louden, Robert (2000) *Kant's Impure Ethics*. Oxford: Oxford University Press.
McDowell, John (1978) 'Are Moral Requirements Hypothetical Imperatives?'. *Proceedings of the Aristotelian Society*, 52, 13–42.
McKenna, Michael (2012) *Conversation and Responsibility*. Oxford: Oxford University Press.
McMullin, Irene (2013) 'Kant on Radical Evil and the Origin of Moral Responsibility'. *Kantian Review*, 18, 49–72.

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press

Michalson, Gordon E. (1990) *Fallen Freedom: Kant on Radical Evil and Moral Regeneration*. Cambridge: Cambridge University Press.
—— (ed.) (2014) *Kant's Religion within the Boundaries of Mere Reason: A Critical Guide*. Cambridge: Cambridge University Press.
Moran, Kate (2014) 'Delusions of Virtue: Kant on Self-Conceit'. *Kantian Review*, 19, 419–47.
Morton, Adam (2004) *On Evil*. New York: Routledge.
Muchnik, Pablo (2010) *Kant's Theory of Evil: An Essay on the Dangers of Self-Love and the Apriroricity of History*. Lanham, MD: Lexington Books.
Munzel, G. Felicitas (1999) *Kant's Conception of Moral Character*. Chicago: University of Chicago Press.
Palmquist, S. (2015) 'What is Kantian *Gesinnung*? On the Priority of Volition over Metaphysics and Psychology in *Religion within the Bounds of Bare Reason*'. *Kantian Review*, 20 (2), 235–64.
Perrett, Roy (2002) 'Evil and Human Nature'. *The Monist*, 85 (2), 304–19.
Russell, Luke (2007) 'Is Evil Action Qualitatively Distinct from Ordinary Wrongdoing?' *Australasian Journal of Philosophy*, 85 (4), 659–77.
—— (2012) 'Evil and Incomprehensibility'. *Midwest Studies in Philosophy*, 36, 62–73.
—— (2014) *Evil: A Philosophical Investigation*. Oxford: Oxford University Press.
Shoemaker, David (2007) 'Moral Address, Moral Responsibility, and the Boundaries of the Moral Community'. *Ethics*, 118, 70–118.
—— (2015) *Responsibility from the Margins*. Oxford: Oxford University Press.
Silber, John (1991) 'Kant at Auschwitz'. In G. Funke and T. M. Seebohm (eds), *Proceedings of the Sixth International Kant Congress* (Washington, DC: Center for Advanced Research in Phenomenology and University Press of America), pp. 177–211.
—— (2012) *Kant's Ethics: The Good, Freedom, and the Will*. Berlin: De Gruyter.
Singer, Marcus (2004) 'The Concept of Evil'. *Philosophy*, 79 (308), 185–214.
Steiner, Hillel (2002) 'Calibrating Evil'. *The Monist*, 85 (2), 183–93.
Straus, Scott (2006) *The Order of Genocide: Race, Power, and War in Rwanda*. Ithaca, NY: Cornell University Press.
Strawson, P. F. (1993) 'Freedom and Resentment'. In John Martin Fischer and Mark Ravizza (eds), *Perspectives on Moral Responsibility* (Ithaca, NY: Cornell University Press), pp. 45–66.
Thomas, Laurence (1993) *Vessels of Evil: American Slavery and the Holocaust*. Philadelphia: Temple University Press.
Vetlesen, Arne Johan (2005) *Evil and Human Agency: Understanding Collective Evildoing*. Cambridge: Cambridge University Press.
Wood, Allen (1999) *Kant's Ethical Thought*. Cambridge: Cambridge University Press.
—— (2014) 'The Evil in Human Nature'. In Gordon Michalson (ed.), *Kant's Religion Within the Boundaries of Mere Reason: A Critical Guide* (Cambridge: Cambridge University Press), pp. 31–57.

https://doi.org/10.1017/S1369415417000140 Published online by Cambridge University Press