

KANT ON THE LIMITS OF HUMAN EVIL

Dr Paul Formosa, University of Queensland

c/o School of HPRC, University of Queensland, St Lucia, QLD, 4072, Australia

p.formosa@uq.edu.au

ABSTRACT

Kant has often been accused of being far too 'optimistic' when it comes to the extremes of evil that humans can perpetrate upon another. In particular, Kant's supposed claim that humans cannot choose evil *qua* evil has struck many people as simply false. Another problem for Kant, or perhaps the same problem in another guise, is his supposed claim that all evil is done for the sake of self-love. While self-love might be a plausible way to explain some instances of evil, it seems to be an implausible way to explain instances where people imprudently act in senselessly destructive and even self-destructive ways. Can Kant handle such extreme cases of moral evil? I shall argue that Kant can handle such cases by: 1) defending Kant's denial of the possibility of a devilish human being; 2) showing how Kant can conceptually account for agents who choose evil *qua* evil; and 3) putting Kant's account of passions to work in order to understand self-destructive evil.

I. INTRODUCTION

In this paper I shall draw upon Immanuel Kant's ethical thought to examine the extremes to which humans can go in the pursuit of evil. But this is likely to seem a misguided project, given that Kant has often been accused of being naively 'optimistic' when it comes to the extremes of evil that humans can perpetrate upon one another. For many of his critics Kant, in Henry Allison's words, "remains very much a child of the *Aufklärung* (or Plato) and, as such, is incapable of recognizing the Dostoevskian depths to which humanity can sink".¹ Dostoevsky's novels are full of characters who feel "a craving to destroy something good".² It is Kant's perceived inability to deal with such cases that has led to the charge that the very worst examples of evil, such as the Holocaust, pose a serious challenge to Kant's overall moral theory. Indeed, John Silber argues that this challenge is so severe that we have grounds for rejecting Kant's entire moral theory, and his account of rationality in particular, because of its inability to deal with such cases of extreme evil.³ The locus of Kant's failure is, according to Silber, to be found in Kant's supposed denial of the human possibility of choosing evil *qua* evil. Another problem for Kant, or perhaps the same problem in another guise, is his supposed claim that all evil is done for the sake of self-love. While self-love might be a plausible way to explain some instances of evil, it seems to be an implausible way to explain instances where people imprudently act in senselessly destructive and even self-destructive ways. It seems *prima facie* grossly misguided to attempt to explain the evil of the Nazi genocide, for example, in terms of self-love. Is Kant able to handle such examples? Or must we explain such cases in terms of the diabolical evil whose possibility Kant supposedly denies and cannot account for? I shall argue that not only can Kant handle such cases, but that his work provides us with a rich and powerful set of conceptual tools for thinking about evil in its very worst manifestations. I support this claim by: 1) defending Kant's denial of the possibility of a devilish human being; 2) showing how Kant can conceptually account for agents who choose evil *qua* evil; and 3) putting Kant's account of passions to work in order to understand self-destructive evil.

II. KANT'S DENIAL OF THE DEVILISH BEING

Kant's discussion of radical evil in his *Religion within the Boundaries of Mere Reason*,⁴ which I have examined elsewhere in depth,⁵ is an attempt to locate the root of all human evil in a corruption at the very root of our power of choice (*Willkür*). For Kant the will (broadly construed) both legislates norms through practical reasoning (*Wille*) and makes executive decisions to adopt maxims in the light of those norms (*Willkür*). As such, Kant rejects two alternative views concerning the source of evil: 1) our sensuous natures, and 2) practical reason or will (*Wille*). It is in the process of rejecting this second view that Kant denies that a human can be a diabolical being. Kant argues that the ground of evil cannot

be placed (2) in a *corruption* of the morally legislative reason, as if reason could extirpate within itself the dignity of the law itself, for this is absolutely impossible. To think of oneself as a freely acting being, yet as exempted from the one law commensurate to such a being (the moral law), would amount to the thought of a cause operating without any law at all (for the determination according to natural law is abolished on account of freedom): and this is a contradiction... A reason exonerated from the moral law, an *evil reason* as it were (an absolutely evil will (ein schlechthin böser Wille)), would ... contain too much, because resistance to the law would itself be thereby elevated to an incentive (for without any incentive the power of choice cannot be determined), and so the subject would be made a *diabolical* [or *devilish*] being (*einem teuflischen Wesen*). – Neither of these two is however applicable to the human being. (6:35)

For Kant a diabolical or devilish being is one who possesses an "evil reason", or equivalently an absolutely evil or fundamentally corrupted will (*Wille*).

Kant denies that any human can possibly be or become a devilish being in this sense. Indeed, Kant argues that it is an a priori conceptual impossibility for a being to possess both freedom and an "evil reason". This claim seems to follow from Kant's so-called Reciprocity Thesis.⁶ As Allison explains: "if, as the Reciprocity Thesis maintains, the moral law necessarily is the law of a free will, in the sense of providing the ultimate norms in terms of which its choice must be justified, then a free rational agent cannot reject the authority of the law without it undermining its own agency".⁷ However, Allison's explanation might seem initially unconvincing because it is far from immediately clear why moral reasons ceasing to have authority for an agent should rob that agent of their very capacity to freely deliberate and act.

A diabolical being suffers from a "*corruption*" of their "morally legislative reason". A diabolical being is not therefore to be confused with a rational amoralist. The former suffers from a corrupted reason (*Wille*), the latter from a corrupted power of choice (*Willkür*).⁸ A rational amoralist is a being who has reason and yet fails to feel strongly the positive 'pull' of morality. Kant discusses such cases in both the *Religion* (6:26-7) and more famously in *Perpetual Peace*.⁹ Kant argues that: "from the fact that a being has reason [it] does not at all follow, simply by virtue of representing its maxims as suited to universal legislation, this reason contains a faculty of determining the power of choice unconditionally, and hence to be 'practical' on its own". (6:26-7) For this reason Kant argues that all humans have a predisposition to (moral) personality, because without such a predisposition we might never find respect for the moral law to be "*of itself a sufficient incentive to the power of choice*". (6:27) While Kant thinks that all humans do have a predisposition to personality (and so can do duty for duty's sake), it is easy enough to imagine a person, or even an entire nation of persons, who no longer find moral motives to be *sufficient* incentives, in and of themselves, for action. Such persons do not lack a predisposition to personality, (in which case they could not *be* moral and thus would lack moral responsibility), but rather fail to *actualise* that predisposition (the reasons for which I return to later in this paper). A rational amoralist can well enough recognise that some maxims are "suited to universal legislation", but *that alone* does not much motivate them to so act. They want to know what is in it for them.

A rational amoralist fails to feel the motivational force of categorical imperatives, whereas a diabolical being cannot even formulate such imperatives because their very reason (*Wille*), and not just their power of choice (*Willkür*), is corrupt at its very core. For this reason a diabolical being, unlike a rationalist amoralist, cannot possibly be a free and responsible agent. For Kant the moral law is a priori in the sense that it is a product of our own rational faculties. The moral law is a law that we each legislate for ourselves through the pure use of our own practical reason. While we can use reason for its own sake in order to generate categorical imperatives, which stipulate objective moral norms, we can also deploy our reason instrumentally to generate hypothetical imperatives, which stipulate norms for achieving our non-universalisable interests and inclinations. But it is the *very same faculty* of reason that generates both hypothetical and categorical imperatives. As such, Kant argues that it is conceptually confused to imagine a faculty of reason that can

both formulate hypothetical imperatives and yet be constitutively unable to formulate categorical imperatives.¹⁰ A being who lacks the capacity to formulate categorical imperatives would therefore also lack the capacity to formulate hypothetical imperatives. Such beings could not therefore reflect on their competing desires and interests, and in the light of these, choose a course of action on the basis of norms (hypothetical *or* categorical) for acting one way or the other. They would not only lack the capacity to act morally, they would also lack the broader capacity to act *rationally*. But then a being who lacked altogether such basic deliberative capacities would indeed lack the capacity for freedom. A corrupted *Wille* therefore undermines the very grounds of responsible agency. No practical reason equals no freedom.

An "evil reason" is not to be confused with an evil *use* of reason. An evil use of reason involves deliberating about the most rational means for achieving an evil outcome, such as the torture of a certain person. But while reason can hypothetically legislate evil means, it cannot categorically legislate evil ends.¹¹ Norms are rationally valid just to the extent that they can be universalised, and are thereby premised on respect for all persons. But an "evil reason" would be the type of *Wille* that is *constitutively* able to *only* legislate 'evil norms' or 'evil laws' as rationally valid and universalisable. But such an "evil reason", such an intrinsically and constitutively 'unreasonable reason', is no reason at all, and cannot therefore form the basis for freedom and responsibility, as made clear in the previous paragraph. A diabolical being, in the sense of a being who is both free and possesses an "evil reason", is an a priori impossibility¹² (at least insofar as we accept Kant's account of agency and freedom in terms of rationality).¹³ This is the "contradiction" that Kant alludes to in the above quoted passage. (6:35)

But even if an "evil reason" is impossible and one's law-giving reason cannot itself categorically legislate 'evil laws', it might still be thought possible that one could be diabolical in the sense of *always* acting immorally *without any struggle or difficulty* whatsoever. Kant calls this possibility a diabolical or wicked *disposition*.¹⁴ Kant argues that:

The depravity of human nature is therefore not to be named *wickedness*, if we take this word in the strict sense, namely as a disposition (a subjective *principle* of maxims) to incorporate evil *qua* evil for incentive into one's maxim (since this is *diabolical*), but should rather be named *perversity* of the heart, and this heart is then called *evil*... An evil heart can coexist with a will which in the abstract is good. (6:37)

Kant, in his account of radical evil, takes human nature to be partly defined by a universal evil disposition to make obedience to the moral law conditional. But as we retain, insofar as we retain our freedom and reason, a will (*Wille*) which in the abstract is good (as opposed to an "evil reason"),¹⁵ our moral state can be no worse than radically evil, which is a state where the moral incentive is present but (at least sometimes) subordinated to other non-moral incentives. However, this does not mean that the moral incentive must explicitly be present in our consciousness each and every time we act. Rather, we must represent, from a practical point of view, the moral incentive as present just insofar as we consider a person to be a free moral agent acting under the constraints of practical reason. We show that the moral incentive is present in us by our very ability to act rationally.

In order to make this position more intuitive, we do well to turn to a parallel line of reasoning whereby Kant argues that:

It is this ... which may distinguish man from a devil, who views himself as governed only by evil itself, and as author of the same, and who therefore, *without struggle or inducement*, engages in no actions other than bad ones.¹⁶

Later in the same lecture series, Kant reiterates this point:

And so too, in his moral behaviour, does man range in thought from the lowest abasement to the highest elevation that he is capable of approaching... Vices, though, like virtues, remain always human, and the maximum of evil, and of good, in devil and angel, is merely an unattainable ideal; a notion of the uttermost degree thinkable that was already to hand before it took on symbolic form under the image of devil and angel.¹⁷

It is important to note here that Kant does not always use the term 'devil' in the same sense. In the earlier quoted passage from the *Religion* (6:35) Kant uses the term to refer to a being with a corrupted *Wille*, whereas in *Perpetual Peace* (8:366) Kant uses the term to refer to a rational amoralist with a corrupted *Willkür*,¹⁸ and in the above passage from the *Lectures on Ethics* Kant uses the term to refer to a being with a diabolical *disposition*. A 'devil' in the first sense, as I argued above, is conceptually impossible, in the second sense both possible and accounted for by Kant, and in the third sense, I shall now argue, while thinkable it just happens to be anthropologically false that humans can be devils in this sense. Just as no human, even the very best, can possess with the angels a holy disposition, equally no human, even the very worst, can possess with the devil a diabolical

disposition. These ideals of pure good and pure evil are beyond us as "rational natural beings".¹⁹ Evil remains always human, in even its very worst manifestations.

As the above passage makes clear, the justification that Kant has in mind for denying the possibility of any human possessing a diabolical disposition must mirror his argument for denying the possibility of any human possessing a holy disposition. For Kant, a holy disposition is a disposition based not on *respect* but on *love* for the law.²⁰ Such a disposition, Kant argues, is beyond us mere humans, for as rational natural beings, our moral duty (at least potentially) requires "self-constraint" in the face of competing natural interests. For this reason, unlike a holy being, we sometimes obey the law at best "*reluctantly* (in the face of opposition from their inclinations)". A holy being is being who never even feels *temptations* to violate the moral law.²¹ As such, no human is holy, for even the most virtuous of us require at least potentially the overcoming of temptation, a challenge a holy being never faces. Our sensuous natures ensure that the moral law remains always an imperative for us that we must choose in the face of potential opposition. The human "moral condition" is always of a "moral disposition *in conflict*".²² For this reason the best humans can obtain is a disposition of respect for the law as their supreme interest.²³ No human is an angel.

But equally, a devilish or diabolical disposition is beyond us for as rational natural beings our evil projects at least potentially require "self-constraint" in the face of both competing rational and natural interests. For this reason, unlike a devilish being, we sometimes violate the "inner moral law *reluctantly*",²⁴ in the face of opposition from our own reason, which we can completely smother only to the extent that we completely smother our very freedom. But a devil never even potentially experiences temptations to violate the moral law for the sake of their self-interest, happiness or natural inclinations. Such a devil is so absolutely committed to doing evil always and only for the sake of doing evil that they cannot even dimly hear the pull of (or even formulate) either a rational interest in morality or a natural interest in achieving happiness or satisfying inclinations.²⁵ As such, no human is diabolical, for even the most vicious of us require at least potentially the overcoming of temptation to act for the sake of either natural or rational interests. This is but the reverse side of Kant's point that our disposition is one always potentially in *conflict*. The devil, no less than the angel, is, unlike us mere humans, beyond a state of even *potential* conflict. No human is a devil

We remain in our conduct always human, stuck between two extremes, neither able to sink so low as to be devils or to raise ourselves so high as to be angels. Human choice is often beset by conflict, struggle and self-constraint in whatever way we choose. Kant argues that humanity's discovery of the "power of choosing for himself a way of life" may have brought "a moment of delight", "but of necessity, anxiety and alarm as to how he was to deal with this newly discovered power quickly followed... He stood, as it were, at the brink of an abyss".²⁶ This angst in the face of the unknown abyss that choice brings with it is indicative of our human condition as rational natural beings, beings torn between competing interests, ends and desires, both rational and sensible, of which we are forced to choose between without being able to fully predict the consequences. As complex rational natural beings ours is a state often marked by anxiety. We can never obtain a state of 'perfection' whereby we are absolutely beyond the possibility of ever doing evil because of our love for the law (as holy beings are) or absolutely beyond the possibility of ever doing good because of our hatred for the law (as devilish beings are).

III. EVIL QUA EVIL

In the previous section I showed that Kant only argues against the possibility of an "evil reason" and a devilish disposition. However, just as a person without a holy disposition can still will to do duty for duty's sake, I shall argue that a person with neither a devilish disposition nor an "evil reason" can still will to violate duty for violation's sake.²⁷ In other words, Kant (or at least the Kantian) can account for the human ability to choose evil *qua* evil.

To support this claim I shall need to show that a being without a fundamentally corrupted *Wille* is *able* to choose evil for evil's sake. Of course, controversy abounds here, depending in part upon one's meta-ethical views. Irit Samet-Porat examines this controversy with admirable clarity by examining the different responses of cognitivists, emotivists, internalists and externalists to the question of whether an agent can choose evil *qua* evil.²⁸ An emotivist, unlike a cognitivist, considers a moral judgment to be an expression of a feeling of approbation and not a truth claim. Thus on an emotivist account, if a "person is attracted by a certain course of action it means that she believes them to be morally good or permissible", even if others consider it wrong or evil.²⁹ The emotivist must

therefore hold that it is conceptually confused to claim that an agent can judge an act to be evil and yet also be attracted to that course of action, for in that case the action could not be evil. As such, there seems to be no way to make sense of the claim that an agent can choose evil *qua* evil on an emotivist account. In contrast, Samet-Porat argues that an externalist about motivation can easily account for the choice of evil *qua* evil. For an externalist there is no necessary link between moral judgements and motivation. That I judge an act to be right does not necessarily imply, for the externalist, that I am attracted to that course of action in anyway. As such, there is no reason why the externalist cannot make conceptual sense of an agent who judges an act to be evil and yet is still attracted to that course of action just because it is evil.

But, of course, Kant is no emotivist or externalist. Samet-Porat argues, as I shall do here, that the Kantian cognitivist-internalist *can* conceptually account for cases of evil *qua* evil. The internalist about motivation, such as Kant, need only claim that moral judgments are necessarily positively motivating. But this need not imply that moral judgments are necessarily *only* positively motivating. I shall argue that on Kant's analysis the same law can both motivate respect (as the internalist demands, thereby proving that my *Wille* is not fundamentally corrupted), insofar as I recognise the universal validity of the law, and also (potentially) hatred and resentment of the law, insofar as my self-conceit is humiliated. However, although it is conceptually coherent (from a Kantian and therefore cognitivist-internalist perspective) to claim that humans can choose evil *qua* evil, I shall also argue in section V that such cases, though possible, are of marginal importance. This is because we can better understand the worst human acts in terms of human passions rather than diabolical perversity.

To choose evil *qua* evil means for Kant to choose evil *immediately*, as a result of the mere consciousness of the moral law,³⁰ and not *mediately*, in order to achieve some antecedent inclination or interest, such as the likes of self-love, envy or malice. As such, to choose to perpetrate evil because of one's hatred for a person, or because of a racist ideology, is *not* to choose evil *immediately* *qua* evil - it is to choose evil *qua* hatred or ideology. Those who advocate the possibility of human's choosing evil *qua* evil often have in mind the examples of the sadist who is immediately drawn to evil because of the pleasure that it affords them, or the racist ideologue who is immediately drawn to evil because of their ideology. However, Kant handles such cases as part of the broader

category of placing the inclinations (including passions – see section V) above the demands of morality. Therefore, to choose evil *qua* evil, and not merely evil *qua* inclination or passion, is to violate the moral law for its own sake, and not for the sake of *any* antecedent interest or inclination, including sadistic or ideological inclinations. This means that the mere consciousness of the moral law must, in itself, be able to provide a *sufficient* incentive to violate the law for its own sake. Is such diabolical perversity possible for beings who have a will which is in the abstract good?

Kant singles out respect as a feeling different in kind from all other feelings. Kant defines all feelings as either intellectual or sensible in origin. As such, “feeling ... is either *pathological* or *moral*. – The former is that feeling which precedes the representation of the law; the latter that which can only follow upon it”.³¹ To do duty for duty’s sake is to act for the sake of an incentive, namely respect, that immediately follows upon, but does not precede, the consciousness of one’s duty. Equally then, to do evil for evil’s sake is to act for the sake of an incentive that immediately follows upon, but does not precede, the consciousness of one’s duty.

For Kant, to recognise the universal validity of a norm is to be at least somewhat and to some extent motivated to act in accord with that norm. It is thus a sign of our practical rationality that the mere consciousness of the moral law *must* engender respect for that law. But need it engender *only* respect? Kant, in the second *Critique*, gives us reasons to think otherwise.³² There he argues that insofar as the moral law thwarts all opposing inclinations, and thereby infringes upon our pursuit of self-love, the recognition of duty “produces a feeling that can be called pain”.³³ The moral law, however, only *thwarts* self-love, because self-love pursued within the boundaries of morality is not opposed by reason. Self-conceit, in contrast, is a hubristic arrogance by which self-love is itself made into a “lawgiving and unconditional practical principle”. The moral law does not merely thwart, but completely “strikes down” such arrogant self-conceit, as reason can allow no other unconditional practical principle but itself. This striking down is a humbling experience that produces “humiliation (intellectual contempt)”. But the moral law still “demands” and “inspires” respect.³⁴ But it is not *only* respect, but also pain and humiliation, that has a moral rather than pathological cause.³⁵

However, Allen Wood argues that: “If the morality of an action represented itself to us similarly as both an incentive for doing it and an incentive for not doing it, then our

moral condition would be incoherent, and once again we could not be considered responsible agents".³⁶ But as Kant makes clear, there is nothing incoherent about the same law inspiring respect in me, insofar as I recognise its universal validity, and also inspiring pain, insofar as my inclinations are thwarted, and humiliation, insofar as I recognise my unjustified self-conceit. Thus it is not for the very same reason that I both respect and am humiliated by the moral law, but it is the very same law that inspires both respect and humiliation.

Given this analysis I shall suggest that there is a specific sense that Kant (or the Kantian) *can* make sense of the possibility that a person might choose evil *qua* evil. Just as we can act for the sake of the positive feeling of respect, so too we can act for the sake of the negative feeling of pain and humiliation. The affront to our self-conceited conception of ourselves, the attack on our pride and the swift rebuke to our arrogance dished out by the consciousness of the moral law can (perhaps) lead to a 'rebellious attitude' of resentment and hatred toward the law itself. To act directly upon this hubristic hatred of the law is to choose evil *qua* evil. It is to choose evil *immediately* and not for the sake of (and indeed in spite of) *any* mediate interest or inclination that precedes the representation of the moral law. Otfried Höffe, in his analysis of Kant's theory of evil, writes of "the possibility of extreme *hubris*, that moral arrogance by which a human being purposely follows maxims that are inherently evil because they contradict the moral law".³⁷ Such unbounded pride in the face of humiliation caused by even the light of our own reason can (perhaps) give vent to a desire to strike down that law, pointlessly and for its own sake, and in complete disregard of one's prior inclinations. But such hubris is not only perverse but self-defeating, for the self-conceited humiliation that gives rise to it fails to achieve the very aims of the self-conceited individual, which is to follow self-love (and thus *preceding* inclinations) unconditionally. Such hubris sacrifices all one's own interests and inclinations in order to spite the law for no other reason than to spite the law as a sort of exaggerated response to the humiliation it imposes.

We can further expand Kant's phenomenology of the consciousness of the moral law by considering the "attraction of forbidden fruit". Georges Bataille argues that "it is always a temptation to knock down a barrier; the forbidden action takes on a significance it lacks ... and invests it with an aura of excitement". People so inspired can "do evil knowingly, precisely because it is evil. It thrills them, which it would not do if it were not evil".³⁸ In

such cases, it is the mere consciousness that the law forbids an action that can *immediately* motivate choosing that action. Hence, the moral law which one respects can also directly motivate, in at least some individuals, the desire to break the law for its own sake, not just from pride or hubris, but also from sheer perversity, simply in order to knock down that which is forbidden just because it is forbidden, whatever the consequences. Further, we find convincing literary accounts of such cases in the work of, among others, Poe and Dostoevsky.³⁹ While Kant does not explicitly consider such phenomenology, there is no reason for the Kantian to deny that some people may indeed be so motivated, and far from being incompatible, this in fact (to the extent that it is accurate) supplements Kant's account of moral psychology. All that Kant needs to maintain is that even such a perverse person must, at some level, also feel respect for the law and must be able to act morally for the sake of that law.

Diabolical perversity is inherently futile and pointless. Its origin is a hatred of the purity of the moral law in comparison with the impurity of our humanity. This can (perhaps) immediately give rise to a rebellious desire to violate that law for its own sake, not in order to achieve any end, and even in defiance of all one's preceding interests and inclinations, simply in order to spite the law (and thus ourselves) out of pride, hubris or perversity. However, Kant does not *explicitly* explore the possibility of such diabolical acts of evil. He does not consider the possibility that someone might inflict evil upon others for its own sake, spurred on by pride in the face of the humiliation that Kant's own analysis of the effect of the moral law on our consciousness so brilliantly brings out. But Kant does leave the conceptual room open for such diabolical cases, even if he does not explicitly consider them. As such, we can find room within Kant's moral philosophy to accommodate, and the conceptual resources to think through, cases of diabolical evil so understood. Of course, this shows us only that such evil is *possible*, and can be accommodated by Kant, and not that people are *in fact* so motivated.⁴⁰ The point is merely that there is no Kantian a priori reason, provided my argument that it makes sense to say that we need not feel *only* respect for the law is successful, why such cases should be conceptually impossible. However, in section V I shall argue that even if such cases are possible, they are not the correct way to primarily think about many large-scale evils, such as the Holocaust, which are exactly the sort of examples that the 'evil for evil's sake' proponent often invokes in order to support their case.

Such cases of diabolical evil are perfectly compatible with Kant's explicit denials of the possibility of an "evil reason" or a devilish disposition. While spiteful hubris or perversity might be a sufficient ground to adopt maxims to pursue evil *qua* evil, this ability cannot turn us into devilish beings. A devilish being has an "evil reason" and therefore lacks agency. A person who chooses evil *qua* evil has agency and a will which in the abstract is good, but simply chooses to make a particularly perverse use of that agency. Further, even if a human being tried to adopt, out of spiteful hubris or perversity, evil *qua* evil as their supreme deponitional interest, this would result, if it were possible, not in a diabolical disposition, but only in a particularly perverse variant of a radically evil disposition. This is because such a choice would not remove the very possibility of conflict from the human moral condition. And this is simply another way of repeating Kant's point – no human, even one who chooses evil *qua* evil, can be a devilish being, just as no human, even one who chooses the good *qua* good, can be a holy being. The human moral condition lies between these inhuman extremes.

IV. SELF-DECEPTION AND EVIL

To further extend this Kantian examination of evil at its worst, we need to examine other sources of evil that Kant accounts for. In this section I shall examine self-deception and the lack of moral cultivation as sources of evil, and in the next section passions and affects. Allison, in arguing in defence of Kant's account of evil, emphasises the important role that self-deception plays in Kant's theory of moral psychology. Allison claims that self-deception, by masking "morally salient features", is the tool whereby deliberate evildoing becomes possible for Kant.⁴¹ Self-deception can allow us to "tinker" with the moral law until we have "fashioned it to suit" our "inclinations and conveniences".⁴² Hannah Arendt similarly argues that for Kant humans are unable to "do evil deliberately" without employing "self-deception". Self-deception is thus, on Arendt's reading of Kant, the "source of all evil," as it is only through self-deception that we are able to perpetrate evil deliberately.⁴³ However, while self-deception is an important moral phenomenon, both Allison and Arendt are wrong to claim that self-deception is a necessary feature of all deliberate evildoing for Kant, or so I shall argue.

To see why, consider the following four examples. Firstly, the case of an agent who makes it his maxim to steal when he can get away with it, but who continues to endorse the norm 'do not steal' as valid. Secondly, the case of an agent who makes it his maxim to 'steal when he can get away with it', and who endorses this as a universally valid norm. Thirdly, the case of someone sometimes steals and sometimes doesn't, not because of any maxim he has adopted, but because he is thoughtlessly directed by outside forces. When he is with one friend he steals, and when he is with another he doesn't, but in either case he never gives the matter much thought. Fourthly, the case of someone who thinks that there are no moral norms, and who steals when he wants to, and has no qualms about doing so.⁴⁴

The first case should be familiar enough to readers of Kant, as it is a case of 'making an exception for oneself'. However, self-deception comes into play here *only* if our thief tries to justify or rationalise his behaviour to himself. He might do so by telling himself that his misfortunate life 'entitles him to it', or that such rules apply 'only in general'. However, self-deception does *not* come into play if our thief is a rational amoralist who intentionally pursues what he himself recognises is an unjustified transgression in order to further his self-interest, which he values more highly than his moral well-being. Kant would see this as a "crime" proper, as it involves an "*intentional* transgression" of the moral law "accompanied by consciousness of its being a transgression".⁴⁵ Such a thief knows he does wrong, and does not attempt to justify or rationalise away his act, or engage in self-deception. His conscience may even trouble him, but he learns to live with that. However, in both cases our thief would continue to think himself wronged if someone stole from him. As such, our thief does not will it to be a universal law that everyone steal when they can. He wants not universal stealing, but to be an exception to that rule. Self-deception enters into the story only if he tries to justify that exception.

The last three cases are more difficult. The second case is one where a person takes stealing to be a valid norm and acts accordingly. No self-deception is required by such a person to engage in wrongful actions, for the simple reason that they do not consider stealing to be a wrongful action. Law-abiding people are considered naive by such a person. The third case is obviously reminiscent of Arendt's account of Nazi bureaucrat Adolf Eichmann in terms of thoughtlessness. I argue elsewhere, in a critical discussion of Allison and Arendt, that self-deception is not in general the right category to think through the

actions of banal perpetrators of evil such as Eichmann.⁴⁶ This is because self-deception, as Allison understands it, presumes a level of thoughtfulness which, if Arendt is correct, banal perpetrators of evil lack. Eichmann and his like do not need to deceive themselves or 'rationalise' their behaviour because they "have never given any thought to the matter".⁴⁷ They are better understood as compartmentalisers (not self-deceivers) whose moral attitudes vary "with their social roles" in ways that they are "quite unaware of".⁴⁸ The fourth case involves someone who does not need to make an exception for himself because he does not recognise any moral norms to be valid. He just 'does what he feels like'. As such, there is no need for him to engage in complicated self-deception.

Of course, it is still possible to read cases two to four as examples of self-deception. In the second case, self-deception might be at play in allowing the person to affirm stealing as a universal maxim, in the third case in allowing the person to be so thoughtless, and in the fourth case in allowing the person to affirm that there are no valid moral norms. The reason why one might think that self-deception is *necessarily* at play is if one agrees with Kant's supposed claim that knowledge of what is right and wrong is a fundamental element of our moral experience. On this view, everyone, even with the humblest intelligence and most rudimentary education, knows as a matter of course what is right and wrong. Hence, as everyone knows automatically what is right and wrong, it follows that, if someone were to get their moral duties mixed up or to take no notice of them (cases two to four), then self-deception *must* be at play.

However, while there is certainly such a "Rousseauian" element in Kant's work,⁴⁹ another strand of Kant's work gives us grounds for holding the more plausible view that the vast majority of humankind do *not* know what is their objective duty in *all* cases. These two strands are perfectly compatible, for the first tells us only that even the humblest intelligence, insofar as they have practical reason and therefore agency, has the *capacity* to legislate categorical imperatives, while the second tells us that the vast majority of humankind do not *exercise* and *cultivate* that capacity.⁵⁰ Kant argues in his paper, *What is Enlightenment*, that a "great a part of humankind" do not think for themselves.⁵¹ To recognise the moral law requires the "resolution and courage" to use one's own faculties of reason and understanding, but most people prefer to stay within the comforts of "minority".

Moral knowledge does not come for nothing. It requires enlightenment and cultivation. In the *Metaphysics of Morals* Kant argues that:

it is incumbent upon him [the moral agent] to enlighten his understanding in the matter of what is or is not a duty... The duty here is only to cultivate one's conscience, to sharpen one's attentiveness to the voice of the inner judge and to use every means to obtain a hearing for it.⁵²

But it is not a matter of course that the 'inner judge' of reason gets a hearing, and if it does not get a hearing, then we are to blame for this, provided that we do not lack a seed of goodness, in which case we would lack practical reason and thus agency itself. As Kant explains: "We no more have special *sense* for what is (morally) good and evil than for *truth*... We have, rather, a *susceptibility* on the part of free choice to be moved by pure practical reason".⁵³ But we have only a "*susceptibility*", and those who do not cultivate and exercise their capacity for pure practical reason can become all but deaf to that susceptibility.⁵⁴ Through misuse, or no use at all, the voice of pure practical reason can wither away to near silence, turning us into moral zombies, the living "morally dead".⁵⁵ But we remain responsible for the silence of our own reason. If we fail to consider the morality of our actions, we fail only because we will to fail. Our reason can be revived, our seed of goodness can sprout anew, and false shoots can be trimmed away. As such, there are always grounds to hope for progress. It is here that Kant's optimism lies. It is not a naive optimism about how well humans will behave, but a reasonable optimism about the heights to which humanity can soar from even the most radically evil of states.

This allows us to better handle the above four cases. The second case is one where the agent has falsely taken a subjective principle to be an objective one. The third and fourth are cases where agents have failed to legislate objective norms at all. We can explain all of these cases as examples of *either* self-deception and rationalising, *or* compartmentalising, *or* a lack of moral cultivation and self-incurred minority – or, more likely, a combination of all of these. Eichmann, for example, always made sure to be covered by superior orders, and this was one way for him to rationalise his behaviour to himself, but he *also* never gave the matter much thought, engaged in a process of compartmentalisation, and failed to exercise his conscience and cultivate his moral understanding. Eichmann's moral failures are thus of multiple origin and interrelated.

All that we require of agents, just insofar as they are agents, is the presence of spontaneity and the seed of goodness that is practical reason. We do not require that agents actually recognise all and only morally valid norms in order to hold them responsible for their actions. All that is needed is that, like all other 'normal' humans, they possess the

ability or *capacity* (i.e. spontaneity and a seed of goodness) through which they *could* come to recognise the validity of moral norms and be thereby somewhat motivated to act morally. But enlightenment is the indispensable path to such a state. Moral reasoning and judgment requires education, cultivation and practise. Indeed, to cultivate one's "moral cast of mind" is for Kant a moral duty, part of implementing the obligatory end of self-perfection.⁵⁶ Those who never bother to thoughtfully consider the morality of their actions allow their conscience to become "weak through habituation".⁵⁷ And those who fail to scrutinise the social mores of the day in the light of their own reason suffer from a self-imposed moral stunted-growth and blindness for which they can be blamed.⁵⁸ If we get in the habit of not asking moral questions in the harsh light of public critical reason, then moral norms can become distorted or silent. Add to this a liberal dose of self-deception, and from such fertile soil evil can easily grow and flourish.

V. PASSIONATE EVIL

In this section I shall turn to Kant's account of passions in order to attempt to understand highly destructive and even self-destructive examples of evil. In section III I examined the sense in which Kant (or the Kantian) can account for the possibility of humans, from spiteful hubris, pride and perversity, choosing evil *qua* evil. But to choose evil *qua* evil is, by definition, utterly pointless, futile and self-defeating. It is to choose evil immediately for no preceding reason whatsoever. For this reason it is a mistake, or so I shall argue, to think of large-scale evils, such as the Holocaust, *primarily* as examples of 'evil *qua* evil'. Emil Fackenheim, for example, attempts to portray the Holocaust as a case of "annihilation for the sake of annihilation, murder for the sake of murder, evil for the sake of evil".⁵⁹ Berel Lang, like Fackenheim, argues that the Nazis "implemented the policy of genocide at least in part *because* it was wrong". Note that Lang qualifies his claim with the 'at least in part' clause. While we cannot absolutely rule out that perhaps mere recognition of wrongdoing did *some* motivational work in *some* cases, I shall argue that it is a mistake to *primarily* (as Lang does) characterise the Nazis' actions in terms of evil *qua* evil. Indeed, as Lang admits, the Nazis also acted upon the "idea" that membership of certain racial groups (i.e. Jews, gypsies, homosexuals etc) "itself suffices to exclude him from the domain of humanity." If the Nazis genuinely believed this justificatory narrative, they not only did

not do evil for evil's sake, they did not even do evil knowing it to be evil. However, Lang identifies the "contradiction, shame and invention" that was characteristic of the Nazi genocide as indicative of the fact "that there was recognition of wrongdoing".⁶⁰ But even if this is correct, it would only show that the Nazis did evil knowing it to be evil, but not necessarily or primarily because it was evil.

In any case, to talk of the 'the Nazis' as if they were a group of identical demons is grossly misleading. Some Nazis were sadists who enjoyed what they did, and others simply obeyed orders without giving the matter any thought. But the thoughtful instigators of the policy of genocide, and Hitler in particular, were not following orders, but issuing them. They decided upon and adopted a policy of genocide. Did they do this just because it was evil? But as we have seen, to choose evil *qua* evil is an utterly futile and pointless act of rebelliousness. This should make us think of random and pointless acts of violence and vandalism, and not of genocide. It should not be surprising therefore that Augustine's famous confessional example of choosing evil for evil's sake is a pointless one-off act of teenage thievery, and not a long-term consciously adopted program of evildoing.⁶¹

To place the radical evil of genocide in the class of 'evil *qua* evil' is misleading on two accounts. First, because it was the likes of demonising hatred, maliciousness, a desire for power and even omnipotence, historical grievances, and a racist ideology, and not the mere fact that genocide is evil, which were the most critical motivational forces behind the Nazis' decision to adopt a policy of genocide. The presence of such inclinations make this a case of evil *qua* hatred, sadism and ideology, and not (or at least not *primarily* or *only*) evil *qua* evil. Second, because a policy of genocide cannot be motivationally sustained on a large-scale and in the long run by a mere pointless act of moral rebelliousness – such a task requires *passion*, energy and commitment. Part of the reason that the Nazis were at least tacitly supported by so many Germans is not because they supported evil or murder per se, but because they swallowed (thoughtlessly perhaps) an ideology that justified and even required those murders. However, without some sort of justificatory narrative those acts could not have garnered (unless we make the mistake of demonising *en masse* the German people) the widespread support that such large-scale evil acts, which need many actors, require. The justificatory narrative itself need not have been particularly plausible, but justificatory narrative there needed to be. Kant's account of self-deception and self-incurred minority helps to explain how such an ideology can spread.

If not in terms of evil *qua* evil, how then can Kant account for the imprudent and self-destructive evil of someone such as Hitler?⁶² Kant has often been thought to be unable to handle such cases because of his supposed conception of all evil-doing in terms of the subordination of morality to self-love (or self-interest). Such a selfish evil-doer, though immoral, is not unreasonable, as they remain open to reasons which show them that evil-doing is not in their self-interest. But the very worst evil-doers are often imprudent and cannot be reasoned with. Arne Johan Vetlesen argues that:

Though not entertained by Kant, what is even more radical than the negation of the good is *the destruction of the good because it is good* – that is, without there being any attempt (be it one involving self-deception and pretence) to *deny* or in any way detract from the goodness of what is deliberately sought destroyed... The lack of any demonstrable *positive spin-off* for the evil-doing agent is a remarkable feature of evil-doing at its most evil – yet frequently overlooked in studies of evil, tending as they do to be wedded to the premise that actions – even evil ones – flow from self-interest.⁶³

Both Sharon Anderson-Gold and Allison attempt to account for the actions of perpetrators of evil in terms of self-interested agents acting in a “competitive social context”.⁶⁴ But not all perpetrators of evil attempt to act in their own or anyone else’s best interests. Some evil persons, as Vetlesen makes clear, do not seek to compete with others for goods, but rather seek to destroy all that is good and do so in imprudent and self-destructive ways. Are such cases beyond Kant? Do we need to appeal to the idea of a devilish disposition or an “evil reason” in order to account for such evil?

To see how Kant can handle such cases, we need to turn to his account of ‘passions’, which he distinguishes from ‘affects’, and apply this to evil. While Kant briefly makes this distinction in the *Metaphysics of Morals*,⁶⁵ I will be drawing here primarily on Kant’s *Anthropology* and *Lectures on Ethics* in order to reinforce and explain this distinction in more depth. Affects, like anger, are feelings which make reflecting difficult or even impossible. In contrast passions, like hatred, are neither “thoughtless” nor “stormy and transitory”, but rather “take root” and “co-exist with rationalising”.⁶⁶ Affects can account for cases of hot-blooded evil-doers who act in the heat of the moment, whereas passions can account for cases of cold and calculating evil-doers who act on the basis of reflection and reason.⁶⁷ When we act from an affect, emotion overpowers our agency (even if this is only because we will it), but when we act from a passion we express our agency and employ our

reason in the pursuit of evil. It is often not rational self-love, or affects, or the diabolical perversity of evil *qua* evil, but rather passions (i.e. evil *qua* passion) that are at the root of evil.

A passion encapsulates a principle to “please one inclination by placing all the rest in the shade or in a dark corner”, and thus a passion blinds an agent to the role of particular inclinations within the larger perspective of “the totality of *all* inclinations”.⁶⁸ As such, passions blind us to the bigger picture, to the detriment of ourselves and others. Passions can, of course, come in various degrees. While all passions involve the neglect of at least some other perhaps very important interests, not all passions are all-encompassing. The stronger the passion, the higher the threshold that must be met before other interests can become operative. If a passion becomes all-encompassing, that threshold is raised so high as to make all other concerns seem by comparison so insignificant as to not be worth pursuing at all. An all-encompassing passion is one that we employ our reason to pursue, single-mindedly, come what may, and to the complete detriment of prudence.

All sorts of inclinations can become passions. The sort of energy and focus that passions inspire in us can even be directed towards evil ends.⁶⁹ Indeed, when an agent allows the likes of hatred, envy, malice, or a racist ideology⁷⁰ to become a deeply-rooted and all-encompassing passion, evil at its very worst (in the sense of ‘most destructive’) can follow. Kant understands envy, for example, as “a propensity to view the well-being of others with distress”.⁷¹ This arises from a “reluctance to see our own well-being overshadowed by another’s”.⁷² An envious person, according to Kant, can endure no one being happy besides themselves and, as such, takes “an immediate satisfaction in the practise” of vice and so can act “like the devil, or the principle of a thoroughly evil being” seeking to “eradicate happiness throughout the world”.⁷³ Malice proper, the opposite of sympathy, is a source of ill will and a desire for things to not go well for others. Malice generates a “malevolent [or sadistic] joy” in the destruction of “what is best in the world as a whole”,⁷⁴ a joy that arises from feeling our own well-being and good conduct more strongly against the misfortune of others. This can lead to a hatred of humankind in general, or a specific group in particular. As such, demonising and malice tend to go hand in hand.⁷⁵

We can add to envy and malice the likes of “malicious glee (*Schadenfreude*)”,⁷⁶ sadism, hatred, fear, jealousy, arrogance, pride, vindictiveness, revengefulness and cruelty

as incentives that beset humanity and can lead to evil, which Kant can and does account for. As such, Kant does not claim that self-love is the *only* motive we have for perpetrating evil acts, although he is often read in this way.⁷⁷ But how can this be reconciled with Kant's position that we act on either principles of self-love or morality? All principles are of two types: those that are categorical, and based on purely universal interests, and those that are hypothetical, and based on non-universal or particular interests. Kant labels the latter *en masse* under the rubric of self-love, as they encapsulate 'selfish' (in the sense of particular or non-universal) interests, even if those interests are not prudential. Hence, Kant does not claim that whenever we act against morality, self-love is the *end* that we seek. As we have seen, envy, malice, ideology and the like can be ends for which we act against morality and sometimes even in spite of our self-interest. But even so, all such cases involve the *principle* of self-love, for such ends can only be chosen in the light of hypothetical (or 'selfish') and never categorical imperatives.⁷⁸ And this follows from the position that pure reason itself is never evil.

Passions, in becoming settled principles and ends for which we act, can even take on "the appearance of [pure] reason".⁷⁹ Such principles, which are necessarily subjective, can appear to the passionate agent as on par with moral or objective principles. Kant calls this "delusion" the "inner practical illusion of taking what is subjective in the motivating cause for objective".⁸⁰ In passions we often stand "convinced" that we have set our own ends, but this is a mere "fantasy" of autonomy, brought on by self-deception, compartmentalisation, thoughtlessness or a lack of moral cultivation.⁸¹ Passions, Kant tells us, "are cancerous sores for pure practical reason"⁸² and, as such, they are as close as we can get to possessing an "evil reason". This explains Kant's view that while we can freely and spontaneously choose evil, we can never *autonomously* choose evil. This does not mean that we cannot choose evil on the basis of reflection, and be able to give reasons for our evil actions in terms of our values.⁸³ Indeed, this is just what a passionately evil agent does. Kant only claims that this is not autonomy, in his formal sense of an action whereby pure reason becomes practical.

The passionately evil agent is one who has adopted an evil end, such as hatred, malice or an evil ideology, as an all-encompassing passion. They pursue their evil passion *come what may*, and this can require self-sacrifice and strength of character.⁸⁴ Instrumental reason will be employed by such agents to formulate plans and strategies for

best achieving their passionately desired evil ends. Such agents cannot be reasoned with, not because they have lost their reason (they haven't), but because their passions have made them deaf to the force of even prudential reasons. In such cases there is no positive spin-off for anyone, the evil agent included, even if everything goes according to plan. It is because of their passionate single-mindedness that such agents can repeatedly pursue evil with no indecisiveness, remorse, or considerations of prudence. In effect, such agents can seem like devils who pursue "intentional transgression" on "principle".⁸⁵ However, passionate evil is not to be confused with evil *qua* evil, for here the inclination or principle (be it hatred or ideology) *precedes* the consciousness of the moral law, and is thus not the immediate consequence of that consciousness. Hence the root of such evil lies in the corruption of our power of choice, and not in the corruption of our pure reason.

The worst evils occur when passions, or evil principles, become not only deeply-rooted and all-encompassing in a few individuals such as Hitler, but also widely (if thoughtlessly) accepted by the masses. While Kant focuses primarily, though not exclusively, on inclinations, such as envy and malice, rather than on ideas, as a source of evil, the twentieth century has illustrated how dangerous the marriage of these two can be. Certain inclinations, such as envy of those better off, as well as the fear and hatred of those who are different, can be widely shared. An ideology that plays upon, incorporates and 'justifies' such inclinations (such as the racist Nazi ideology) can expect to find fertile soil amongst humanity, especially in times of crises, when people tend to look for security within exclusionary group identities.⁸⁶ Of course, only passionately evil agents will be thoughtfully convinced by such evil principles, and will pursue them for their own sake, come what may. But where a justificatory narrative (even if not a very good one) is offered for an evil agenda, especially one that plays upon tendencies to demonise others, then the banal masses (as Arendt calls them) can come to thoughtlessly support such evil policies. A passionate and demonising hatred, coupled with an ideology, can spread like a virus along the surface of the masses (the crew of any political ship). When this happens, we can be sure that evil will soon follow in its wake.

This can lead to the wide-spread and counter-prudential pursuit of the destruction of all that is good, for its own sake, come what may. As such, passions can morph, in their extreme, into the maximum of viciousness. For Kant virtue is proportional to the resistance overcome. It is all very well and good to help someone (where it is my duty to do so) when

I am already sympathetically inclined towards them. But to do so even when I am disinclined to and when it is conducive to neither my happiness nor self-interest, is more virtuous for Kant because a greater strength of will and firmness of character is required to overcome strong obstacles. The maximum of virtuousness is to act on moral principle in the face of enormous opposition from interests and inclinations alike. Similarly, vice is also proportional to the resistance overcome. It is understandable (though reprehensible) to wrong someone when I am very strongly inclined to do so (they may be my hated enemy) and when it is conducive to both my happiness and self-interest. The maximum of viciousness is to perpetrate evil 'on principle' and in the face of enormous opposition from interests and inclinations alike. The agent who perpetrates evil because they are passionately committed to a hate-filled ideology (either as a result of self-deception or self-incurred minority) even though all their other inclinations (such as sympathy) and self-interest point in the other direction, is such a maximally vicious agent.

To illustrate this, we do well to turn to Herman Melville's Ahab. Ahab's passion is the destruction of Moby-Dick, and he is "ready to sacrifice all mortal interests to that one passion".⁸⁷ But while Ahab abandons prudence, he does not abandon reason – as Ahab himself puts it, "all my means are sane, my motive and my object mad".⁸⁸ Ahab's fateful third and final attempt to kill Moby-Dick, after two disastrous attempts, is so horrifying precisely because, as Ahab himself realises at Starbuck's promptings, nothing speaks in favour of the continuing pursuit – nothing, that is, except Ahab's blind and passionate need for vengeance on Moby-Dick. This passion drives Ahab to the destruction of himself and his entire crew, and good Starbuck, who foresees all this, is impotent in his feeble righteousness to stand up to such a great storm, a storm that has infected the entire crew with its fever.

There are, perhaps, certain similarities to be drawn between charismatic Ahab, who part inspires and part tyrannises his crew into league with his self-destructive mission, and the likes of Hitler leading Germany to its downfall. Like Ahab, Hitler embarked on a hate-filled quest of destruction, and in the end came to ignore prudence and self-interest, thereby courting his own and his country's self-destruction. For example, as Arendt notes, Hitler ignored the prudent measure of using scarce trains, near the end of the war, to transport supplies to troops, rather than to transport Jews to their death.⁸⁹ But this is expressive, not so much of Hitler's complete irrationality, but rather of the fact that Hitler's

overriding passion had become the destruction of the Jews and not the winning of the war or self-preservation. Some of Hitler's crew, no doubt, were infected by his passionate hatred and malice, spread through a racist ideology. This surely led some Nazis to perpetrate evil on 'principle' and in the face of their other inclinations (they may even have felt sympathy towards their victims) and self-interest. Indeed, it was just this ability that Himmler and the SS prided themselves on. One needed to be 'strong' and 'overcome' such 'weak' sympathetic feelings, and sacrifice self-interest to the historic and grandiose task of establishing the thousand year Reich through the genocide of 'inferior' races.⁹⁰ This is evil at its most vicious – the passionate destruction of goodness for the sake of 'principle', against prudence and all interest and inclination. But, alas, such senseless and horrifying carnage is not beyond the feats of humanity.⁹¹ Who needs devils when we have humans?

VI. CONCLUSION

Kant, far from being naively 'optimistic' about the limits to which human evil can go, provides rather an enormously rich, deep and powerful set of conceptual resources for thinking about evil in its various and worst manifestations. I have sought to defend Kant's core claim that no human can possess an "evil reason" or an absolutely evil will (*Wille*). Further, while I have suggested that the Kantian can account for the human ability to choose evil *qua* evil as at least a conceptual possibility, I have also argued that we can often better understand self-destructive and imprudent evil, not in terms of evil *qua* self-interest or evil *qua* evil, but rather in terms of evil *qua* passions. However, all humans, insofar as they remain agents, retain a seed of goodness in the form of pure practical reason and a susceptibility (which may be very weak or very strong) to a moral feeling of respect when recognising the validity of the moral law. Beyond this, the sorts of evils that humans can perpetrate are almost infinite. Imagination and passion can find an outlet in evil, no less than in beauty and goodness. The depths to which humanity can sink are almost bottomless. But no matter how far we sink, as long as we retain our agency, and thus the capacity for pure reason to practically determine our actions, the hope for progress need not be completely lost. Humanity has it in it to approach the perfection of angels no less than the depravity of devils.

Finally, while this may be all very good as an immanent defence of Kant, some critics may still wish to ask: what is in this account for the non-Kantian? Lots, I hope to have shown. Kant's account of passions as particular inclinations, whether based in feelings (including those of hatred, envy and the like) or ideas (in particular in the form of ideology), that consistently and irrationally dominate over others, is an important insight into evil that is not tied in any strong way to Kant's overall moral theory. Likewise, the discussion of what constitutes evil *qua* evil, and the possibility and actuality thereof, is of a sufficiently general nature to be of interest to the non-Kantian, although of course emotivists (and perhaps externalists) will not be able to endorse this account. Similarly, Kant's denial of the diabolical disposition is based on the fairly neutral moral premise that the human moral condition is one necessarily defined by potential conflict. However, the most characteristically Kantian claim is embedded in the denial of an "evil reason". This claim will probably make little sense to those who see reason as a mere slave of the passions and thus unable to autonomously legislate moral laws.⁹² Further, those who deny the Kantian claim that agency and responsibility are to be understood in terms of rationality may (perhaps) be able to countenance the possibility of a being with a genuinely "evil reason". However, the proponent of such a view must defend the (arguably implausible) claim that an agent who, even though they possess an absolutely corrupted reason (*Wille*) and cannot therefore even *possibly* be rational, is still a moral agent who warrants responsibility for their actions. In any case, I have shown that Kant's denial of the conceptual coherency of such a case does not diminish the explanatory power of his account of evil. As such, far from being a weak spot in Kant's moral theory, Kant's account of evil-doing has much to teach Kantians and non-Kantians alike.⁹³

¹ See Henry E Allison, *Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy* (Cambridge: Cambridge University Press, 1996), 176.

² See, for example, Fyodor Dostoevsky, *The Brothers Karamazov*, trans. Constance Garnett (Chicago: Encyclopaedia Britannica, 1952), Book XI, chap III.

³ See John R Silber, "Kant at Auschwitz," in *Proceedings: Sixth International Kant Congress*, ed. Funke and Seebohm (Washington, D.C.: The Center for Advanced Research in Phenomenology, 1985), 194 and John R Silber, "The Ethical Significance of Kant's Religion," in *Religion within the Limits of Reason Alone* (New York: Harper Torchbooks, 1960).

⁴ Immanuel Kant, "Religion within the Boundaries of Mere Reason," in *Religion and Rational Theology*, ed. Allen Wood and George Di Giovanni (Cambridge: Cambridge University Press, 1996), 6:34-7. All in-text page references are to this book.

⁵ See Paul Formosa, "Kant on the Radical Evil of Human Nature," *The Philosophical Forum* 38, no. 3 (2007).

⁶ See Chapter 11 on the 'The Reciprocity Thesis' in Henry E Allison, *Kant's Theory of Freedom* (Cambridge: Cambridge University Press, 1990), 201-13.

⁷ Allison (1996), 175.

⁸ Although Kant's radical evil thesis implies that we all (at least initially) have a corrupted power of choice. As such, the rational amoralist differs from the rest of us only in degree and not in kind.

⁹ Where Kant talks of a "a nation of devils" (*ein Volk von Teufeln*) who have "understanding" - Immanuel Kant, "Toward Perpetual Peace," in *Practical Philosophy*, ed. Mary J Gregor (Cambridge: Cambridge University Press, 1996), 8:366. Kant's point here is that we can justify principles of justice both in terms of rational self-interest and in terms of morality.

¹⁰ To repeat: a rational amoralist *can* formulate categorical imperatives (although they might not) and they *can* recognise the validity of such norms (although again they might not - see section IV), even though they find hypothetical imperatives to be more motivating than categorical ones (for reasons explored in sections IV and V).

¹¹ Of course, it is sometimes claimed that reason can *only* calculate means and never set ends - see Jürgen Habermas, "Discourse Ethics: Notes on a Program of Philosophical Justification," in *Moral Consciousness and Communicative Action* (Cambridge: MIT Press, 1990), 43.

¹² For a similar claim, see Matthew Caswell, "Kant's Conception of the Highest Good, the *Gesinnung*, and the Theory of Radical Evil," *Kant-Studien* 97 (2006), 201.

¹³ But to reject this account is to endorse the (arguably implausible) position that the mere *potential* to act rationally is not a necessary precondition of agency.

¹⁴ A disposition (*Gesinnung*) is the supreme maxim that is the subjective ground of all other maxims. (6:25)

¹⁵ Elsewhere Kant calls this "the idea of devilish evil" or a being with "no seed of good at all, not even a good will" - Immanuel Kant, *Lectures on Ethics*, ed. Peter Heath and J B Schneewind, trans. Peter Heath (Cambridge: Cambridge University Press, 1997), 27:316-17. From the Collins notes of 1784-5.

¹⁶ Kant (1997), 27:572. Italics mine, from the Vigilantius notes, 1793.

¹⁷ Kant (1997), 27:691.

¹⁸ See endnote nine.

¹⁹ Kant, "The Metaphysics of Morals," in *Practical Philosophy*, ed. Mary J Gregor (Cambridge: Cambridge University Press, 1996), 6:379.

²⁰ See Immanuel Kant, "Critique of Practical Reason," in *Practical Philosophy*, ed. Mary J Gregor (Cambridge: Cambridge University Press, 1996), 5:83-6.

²¹ In a state of holiness "the law would finally cease to be a command for us, since we could never be tempted to be unfaithful to it" - *Ibid.*, 5:82.

²² See *Ibid.*, 5:32; 5:84.

²³ However, Kant is somewhat ambivalent about this possibility (of a disposition of reverence for duty), suggesting that perhaps we need the grace of God to overcome our radical evil and achieve such a disposition. In any case,

Kant is unequivocal that we have a duty to "strive" for, but cannot ever achieve, a holy disposition - see *Ibid.*, 5:83.

²⁴ See Kant, "The Metaphysics of Morals," (1996) 6:379-80.

²⁵ Again, this reinforces the point that the rational amoralist does not have a diabolical disposition.

²⁶ Immanuel Kant, "Conjectural Beginning of Human History," in *On History*, ed. Lewis White Beck (Indianapolis: Bobbs-Merrill, 1963), 56.

²⁷ Samet-Porat also emphasises the importance of keeping Kant's denial of a diabolical being separate from the question of whether an agent without a fundamentally corrupted *Wille* can choose evil *qua* evil - see Irit Samet-Porat, "Satanic Motivations," *Journal of Value Inquiry* 41 (2007), 88-9.

²⁸ See *Ibid.*, 77-97.

²⁹ *Ibid.*, 81.

³⁰ As Allison notes, following Lewis White Beck, it is the "consciousness of the law, not the law itself, that functions as the actual moral motive" - Allison (1990), 122.

³¹ Kant, "The Metaphysics of Morals," (1996), 6:399.

³² See 'On the incentives of pure practical reason', in Kant, "Critique of Practical Reason," (1996), 5:72-89.

³³ *Ibid.*, 5:73.

³⁴ Kant, "Critique of Practical Reason," (1996), 5:80.

³⁵ Kant argues that: "As the effect of consciousness of the moral law, and consequently in relation to an intelligible cause, namely the subject of pure practical reason as the supreme lawgiver, this feeling of a rational subject affected by inclinations is indeed called humiliation (intellectual contempt); but in relation to its positive ground, the law, it is at the same time called respect for the law; there is indeed no feeling for this law, but inasmuch as it moves resistance out of the way, in the judgment of reason this removal of a hindrance is esteemed equivalent to a positive furthering of its causality. Because of this, this feeling can now be called a feeling of respect for the moral law, while on both grounds together it can be called a *moral feeling*." See *Ibid.*, 5:75. Kant with his 'on both grounds together' clause attempts to exclude humiliation, which is not pathologically caused, from being called a 'moral feeling'. However, the *Metaphysics of Morals* definition of a moral feeling, quoted above, in terms of a feeling that follows upon, rather than precedes, the representation of the law, makes no such distinction and so allows humiliation to be called a 'moral feeling'. In any case, moral or not, it is not a pathological feeling.

³⁶ Allen Wood, *Kant's Ethical Thought* (Cambridge: Cambridge University Press, 1999), 285.

³⁷ Otfried Höffe, *Kant's Cosmopolitan Theory of Law and Peace*, trans. Alexandra Newton (Cambridge: Cambridge University Press, 2006), 71. Bernstein similarly argues that: "if recognition of the moral law can serve as an incentive to act morally, there can always be a counter-incentive. We can choose ... to be devilish, we can choose to defy the moral law" - Richard J Bernstein, *Radical Evil: A Philosophical Interrogation* (Cambridge: Blackwell Publishers, 2002), 41-2. However, both Höffe and Bernstein fail to draw upon the second *Critique* to show that Kant already accounts for such motivations.

³⁸ See Georges Bataille, *Erotism: Death and Sensuality*, trans. Mary Dalwood (San Francisco: City Lights, 1986), 48, 72, 116.

³⁹ Poe writes of the "spirit of perverseness" which is a "primitive" human impulse to make oneself "wretched beyond the wretchedness of mere humanity" - see Edgar Allan Poe, "The Black Cat," in *Tales of Mystery and*

Imagination, ed. Graham Clarke (London: Everyman's Library, 2002), 566, 69. We encounter a similar 'spirit' in some of the characters that inhabit Dostoevsky's novels and stories. The narrator of *Notes from the Underground*, for example, writes in this 'spirit' that: "Whether it is good or bad, it is sometimes very pleasant to smash things" - Fyodor Dostoevsky, "Notes from the Underground," in *Great Short Works of Fyodor Dostoevsky* (New York: Perennial Classics, 2004), 291.

⁴⁰ As Samet-Porat notes, some philosophers admit the conceptual coherency of cases of evil *qua* evil, but deny that there have *actually* been any such cases of evil - see Samet-Porat (2007), 84. However, once we can account for the conceptual coherency of such cases, it seems to me to be very likely, and indeed almost undeniable, that there are cases of agents who choose evil *qua* evil.

⁴¹ Allison claims that: "rational agents can of course adopt foolish or immoral maxims, but they cannot adopt maxims without taking them to be in some sense justified (although this may very well rest on self-deception)" - Allison (1990), 91. See also Allison's discussion of Eichmann in Allison (1996), 169-82.

⁴² Kant (1997), 27:464-5.

⁴³ See Hannah Arendt, "Basic Moral Propositions," in *The Library of Congress, The Hannah Arendt Papers* (1966).

⁴⁴ Kant calls this "precarious state of the human mind," which "takes from moral laws all their force as incentives to the heart, and over time all their authority", "libertinism, i.e. the principle of recognising no duty at all" - Immanuel Kant, "What Does It Mean to Orient Oneself in Thinking?," in *Religion and Rational Theology*, ed. Allen Wood and George Di Giovanni (Cambridge: Cambridge University Press, 1996), 8:146.

⁴⁵ Kant, "The Metaphysics of Morals," (1996), 6:223.

⁴⁶ See Paul Formosa, "Moral Responsibility for Banal Evil," *Journal of Social Philosophy* 37, no. 4 (2006), 513.

⁴⁷ Hannah Arendt, "Some Questions of Moral Philosophy," in *Responsibility and Judgment*, ed. Jerome Kohn (New York: Schocken Books, 2003), 94-5.

⁴⁸ On compartmentalisation, see Alasdair MacIntyre, *Ethics and Politics: Selected Essays, Volume 2* (New York: Cambridge University Press, 2006), 196-200.

⁴⁹ As Allison notes: "in spite of his adherence to the Rousseauian principle that even the humblest intelligence is capable of distinguishing right from wrong, Kant does at times acknowledge the possibility of being mistaken about what is objectively one's duty" - Allison (1996), 180.

⁵⁰ As Timmerman notes, on Kant's account we can make an "objective error of judgment" in regard to moral matters, as we do not always reason correctly, or even reason at all - see Jens Timmermann, "Kant on Conscience, 'Indirect' Duty, and Moral Error," *International Philosophical Quarterly* 46, no. 3 (2006), 303.

⁵¹ See Immanuel Kant, "An Answer to the Question: What Is Enlightenment?," in *Practical Philosophy*, ed. Mary J Gregor (Cambridge: Cambridge University Press, 1996), 8:35.

⁵² Kant, "The Metaphysics of Morals," (1996), 6:401. See also Timmermann (2006).

⁵³ Kant, "The Metaphysics of Morals," (1996), 6:400.

⁵⁴ Of course, it follows from section II that we cannot become *completely* and *irreversibly* deaf to pure reason, or we would no longer be moral agents who can reasonably be held responsible for our actions.

⁵⁵ Ibid. See also Kant, "Critique of Practical Reason," (1996), 5:61.

⁵⁶ Kant, "The Metaphysics of Morals," (1996), 6:386-7.

⁵⁷ Kant (1997), 27:356.

⁵⁸ Kant argues that when "custom ... is raised to the dignity of a law", what results is "a tyranny of popular mores [which] would be contrary to his duty to himself" - Kant, "The Metaphysics of Morals," (1996), 4:464.

⁵⁹ Quoted in Emmanuel Levinas, "Useless Suffering," in *The Provocation of Levinas: Rethinking the Other*, ed. Robert Bernasconi and David Wood (London: Routledge, 1988), 163.

⁶⁰ Berel Lang, *Act and Idea in the Nazi Genocide* (Syracuse, New York: Syracuse University Press, 2003) 19-32.

⁶¹ See Book II from Saint Augustine, *Confessions*, trans. R S Pine-Coffin (Middlesex: Penguin Books, 1961).

⁶² Silber argues that "Kant's theory ... cannot illuminate the conduct of a Hitler" - Silber (1985), 194.

⁶³ Arne Johan Vetlesen, *Evil and Human Agency: Understanding Collective Evildoing* (Cambridge: Cambridge University Press, 2005), 128.

⁶⁴ Sharon Anderson-Gold, "Kant's Rejection of Devilishness: The Limits of Human Volition," *Idealistic Studies* 14, no. 1 (1984), 36.

⁶⁵ Kant, "The Metaphysics of Morals," (1996), 6:408.

⁶⁶ Immanuel Kant, *Anthropology from a Pragmatic Point of View*, ed. Robert B Louden, trans. Robert B Louden (New York: Cambridge University Press, 2006), 7:265.

⁶⁷ Kant's claim that "great crimes are paroxysm", something like a "fit of madness" whose basis is the "force of inclinations that *weaken* reason, which proves no strength of soul" is, as a general claim, misleading - Kant, "The Metaphysics of Morals," (1996), 6:384. For this makes sense only of cases of affect and not passion. To stick firmly and resolutely to one's passionate end does show a certain depth and strength of character, in Kant's sense of 'character' as one who sticks firmly to principles, even evil ones. For a defence of the claim that evildoers can have character, see Patrick Frierson, "Character and Evil in Kant's Moral Anthropology," *Journal of the History of Philosophy* 44, no. 4 (2006). Of course, other evildoers, like Eichmann, lack character.

⁶⁸ Kant (2006), 7:266.

⁶⁹ "The calm with which one gives oneself up to it [a passion] permits reflection and allows the mind to form principles upon it and so, if inclination lights upon something contrary to the law, to brood upon it, to get it rooted deeply, and so to take up what is evil (as something premeditated) into its maxim." - *Ibid.*, 7:266, 71.

⁷⁰ It is not only sensible inclinations but also ideas that can become passions. Kant considers the example of hatred of injustice turning into hatred against "*him who is unjust to us*" and writes: "Since this inclination (to pursue and destroy) *is based on an idea* [my italics] ... it transforms the desire for justice against the offender into the passion for retaliation, which is often violent to the point of madness, leading a man to expose himself to ruin if only his enemy does not escape it" - *Ibid.*, 7:271.

⁷¹ Kant, "The Metaphysics of Morals," (1996), 6:458.

⁷² *Ibid.*, 6:458-9.

⁷³ Kant (1997), 27:440. Note that Kant only says "*like*" the devil.

⁷⁴ Kant, "The Metaphysics of Morals," (1996), 6:460.

⁷⁵ On the connection between demonising and evil see my paper, Paul Formosa, "The Problems with Evil," *Contemporary Political Theory* forthcoming (2008).

⁷⁶ Kant (1997), 27:440.

⁷⁷ Arendt reads Kant in this way - see Hannah Arendt, *The Origins of Totalitarianism* (London: George Allen and Unwin, 1967), 459. Korsgaard similarly argues that "Kant was mistaken in focusing exclusively on the principle of

self-love... Evil *may* [take this form], but it takes other forms as well" - Christine M Korsgaard, *The Sources of Normativity*, ed. Onora O'Neill (Cambridge: Cambridge University Press, 1996), 250. Bernstein makes a similar point: "why not recognize that there are other incentives [for choosing evil] that are not easily assimilated to 'self-love'" - Bernstein (2002), 42.

⁷⁸ See Allison (1990), 103, and Wood (1999), 400.

⁷⁹ Kant (2006), 7:266, 71.

⁸⁰ *Ibid.*, 7:274.

⁸¹ *Ibid.*, 7:275.

⁸² *Ibid.*, 7:266, 71.

⁸³ And this is all that we usually mean by 'autonomy', and so in this sense we can 'autonomously' choose evil - for this sense of autonomy see, for example, Merle Spriggs, "Can We Help Addicts Become More Autonomous? Inside the Mind of an Addict," *Bioethics* 17, no. 5-6 (2003), 545.

⁸⁴ Silber argues against Kant's account of evil because evildoers like Hitler and Ahab do not languish "in the impotency of personality demanded by Kant's conception of freedom" - see Silber (1960), cxxix. But Kant's account demands no such thing. While evildoers may lack *moral* personality, it does not follow that they necessarily lack a strong physical personality or a powerful and charismatic character - see Kant, "The Metaphysics of Morals," (1996), 6:223. While some leaders, like Hitler, lack moral personality, others, like Ghandi or Mandela, do not. But in either case the ability to be a strong leader requires above all else the right sort of physical personality, character, oratory skills and so on. Indeed, far from denying such cases, Kant's account of passions can positively account for them.

⁸⁵ Kant, "The Metaphysics of Morals," (1996), 6:390.

⁸⁶ See Ervin Staub, *The Psychology of Good and Evil* (Cambridge: Cambridge University Press, 2003), 419.

⁸⁷ Herman Melville, *Moby-Dick or the Whale* (Hertfordshire: Wordsworth Classics, 2002), 176.

⁸⁸ See chapter 41, in *Ibid.*, 154.

⁸⁹ Hannah Arendt, "The Concentration Camps," *Partisan Review* XV, no. 7 (1948), 749.

⁹⁰ See Hannah Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil* (New York: Penguin Books, 1965), 105.

⁹¹ Kant does consider the possibility of a "war of extermination (*bellum internecinum*)", although only as a case of illegitimate warfare - Kant, "The Metaphysics of Morals," (1996), 6:347. But we need only consider Kant's long litany of charges against humanity in the *Religion* (6:33-34) to assure ourselves that Kant was certainly not too 'optimistic' to consider even the very worst crimes, genocide included, as beyond humanity.

⁹² See Samet-Porat (2007), 83.

⁹³ Thanks to Marguerite La Caze, Arne Johan Vetlesen, the journal's editor, Michael R. DePaul, the journal's two anonymous referees, and to all those who offered feedback on the presentation of this paper at the University of Queensland and at the 2007 Australasian Association of Philosophy Conference.