

## Kant on the Radical Evil of Human Nature

### 1. Introduction

In his book *Religion within the boundaries of mere reason*, Kant presents his thesis that human nature is 'radically evil'.<sup>1</sup> To be radically evil is to have a corrupted moral orientation or, equivalently, an evil disposition. However, the very coherency of Kant's radical evil thesis has often been questioned, as has the nature of the argument Kant supposedly offers for this thesis. Kant's argument for radical evil consists primarily of two parts: an evil disposition derivation, where Kant argues that from a single evil maxim one can infer an evil disposition, and the universality claim, where Kant argues that *all* humans have an evil disposition. The first but not second of these arguments succeeds. Even so, radical evil is likely to be very widespread, if not universally so, amongst humanity. As such, Kant's thesis deserves to be taken seriously by *any* moral and political theory. It deserves to be taken seriously, or so I shall argue, because it paints an eminently plausible picture of the human moral condition.

### 2. Kant's radical evil thesis

It is well known that Kant clarifies and expands his account of moral psychology in the *Religion* by arguing that the faculty of volition is composed of two unified but distinct parts: will (*Wille*) and power of choice (*Willkür*). This distinction is based on the fact that 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*). A maxim is not something one just has, like a certain feeling, but is something that is freely and spontaneously *adopted*, through the 'incorporating' of incentives, by one's power of choice.<sup>2</sup> Our power of choice can freely and spontaneously choose either good or evil maxims, thereby acting autonomously or heteronomously respectively.<sup>3</sup> In the

---

<sup>1</sup> All page references in the text and notes, unless otherwise indicated, are to 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).

<sup>2</sup> See Henry E Allison, *Kant's Theory of Freedom* (Cambridge: Cambridge University Press, 1990) 129-36.

<sup>3</sup> As is often noted, this allows Kant to reject the common complaint, first made by Henry Sidgwick, that he is unable to account for our ability to freely adopt evil maxims – see the discussion in Matthew Caswell, "The Value of Humanity and Kant's Conception of Evil," *Journal of the History of Philosophy* 44, no. 4 (2006): 645.

former case the moral law is the incentive for which the maxim is adopted and in the latter case a prudential law or sensuous impulse is the incentive for which the maxim is adopted. But, as Kant explains:

Whether the human being is good or evil, must not lie in the difference between the incentives [of self-love and morality] that he incorporates into his 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. (6:36)

An important consequence of this is that the ground of evil cannot be located in either “the sensuous nature of the human being” itself, nor in the natural inclinations which arise from it. (6:35) This is because evil only arises when we freely *adopt* a maxim that subordinates the incentives of morality to other non-moral interests. But why do we do this? Kant’s radical evil thesis is an attempt to answer this question by locating the root of all evil maxims.

However, in order to correctly understand Kant’s radical evil thesis, we need to first understand the way that Kant uses three key technical terms – namely, ‘predisposition’ (*Anlage*), ‘propensity’ (*Hang*) and ‘disposition’ (*Gesinnung*).<sup>4</sup> Kant defines three predispositions to the good, each of which are not only negatively good, in the sense that it does not resist the moral law, but are also positively good, in the sense that compliance is required, (6:28) although only within the constraints of the moral law. Kant derives these three predispositions to the good by thinking of persons as animal, human and moral beings respectively. The first predisposition to the good is to *animality*. This predisposition is based on what Kant calls mechanical or physical self-love, by which he means love for which reason is not required. Specifically, this takes three forms: self-preservation, propagation of the species through the sexual drive, and community with other beings through the social drive. The predisposition to animality, if misused, can lead to the “savagery of nature” inherent in the “*bestial vices of gluttony, lust and wild lawlessness* (in relation to other human beings)”. (6:26-7) The second predisposition is to *humanity* and is based on a form of self-love which is both physical and rational, as it involves comparison, for which reason is required. Out of this self-love originates a desire

---

But it remains the case that we do not act autonomously, in Kant’s technical sense of the term, when we adopt evil maxims, even if we do so freely and spontaneously after much careful deliberation.

<sup>4</sup> It is important to note that Kant’s terms ‘*Anlage*’ and ‘*Gesinnung*’ are not as closely related as their English translations, ‘predisposition’ and ‘disposition’, might suggest.

for comparison of one's own social circumstances with others, which leads to competition and the progress of humanity. However, if misused, this predisposition leads to what Kant calls the "vices of *culture*", which, in their "extreme degrees of malignancy" lead to the "*diabolical vices*" of "*envy, ingratitude, [and] joy in others' misfortune*". (6:27) The third predisposition is to *personality*, which concerns the human being solely intellectually (or rationally) and is defined as the "susceptibility to respect the moral law *as of itself a sufficient incentive to the power of choice*". (6:27) This last predisposition points to the 'fact of reason' that we find the moral law in itself motivating.<sup>5</sup>

The first predisposition is necessary for humans to exist in nature, the second predisposition forms the basis of the motivation to progress from nature to culture, while the last predisposition guarantees the normative thrust of morality and forms the motivation to progress towards a moral culture. These predispositions to the good, as a whole, form for Kant a *telos* whose end is the achievement of a state of perpetual peace in which cosmopolitan moral cultures can flourish.

However, Kant also argues that humanity has a freely chosen 'propensity' to adopt evil maxims. Unfortunately, Kant initially gives two different definitions of a propensity which are neither equivalent nor even clearly compatible. Kant writes:

By *propensity (propensio)* I understand the subjective ground of the possibility of an inclination...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 presented as not being such: it can rather be thought of (if it is good) as *acquired*, or (if evil) as *brought* by the human *upon* himself. (6:29)

But then, in a way that might at first confuse the reader, Kant adds in a footnote to the first of the above-quoted sentences: "*Propensity* is actually only the *predisposition* to desire an enjoyment which, when the subject has experienced it, arouses *inclination* to it". (6:29) On the first of these definitions a propensity is different from a predisposition, in that it is brought upon oneself or acquired, but is like a predisposition in that it can be innate, even though apparently it need not be represented as such. On the second definition, a propensity just *is* a predisposition, though perhaps only of a special kind. Even worse, Kant makes the second definition seem to be an elucidation of the first, when in fact the two definitions clearly differ.

---

<sup>5</sup> Kant claims that all three predispositions "are original for they belong to the possibility of humanity," (6:28) however it is difficult to see how the third predisposition, which concerns not the possibility of *humanity* but *personality*, can be covered by such a claim.

Thankfully, these difficulties are easily enough resolved later in the text when Kant argues that a propensity can be either moral (this is the first definition), and thereby pertains to us as free beings, or physical (this is the second definition), and thereby pertains to us as natural beings. Therefore, as a physical propensity pertains to us as mere natural beings, it does not have its roots in our freedom, and so it cannot be represented as 'brought upon ourselves'. In this case a propensity, as the second definition above makes clear, just is a predisposition, in that it is part of our unchosen nature. As such, the alcoholic who has a physical propensity to alcohol is not responsible for that physical propensity, but only for the use they make of it. In contrast, a moral propensity has its roots in our freedom. It is a maxim which is the subjective determining ground of all other maxims. Thus a moral propensity is completely unlike a predisposition in that it is a maxim and not an unchosen part of our nature.

To make matters even more complex, a moral propensity is in fact equivalent to what Kant refers to in both the *Religion* text and elsewhere as a 'disposition'. Both terms refer to the supreme maxim that is the subjective ground of all other maxims.<sup>6</sup> Why then does Kant introduce another term? One reason why it might be useful to have two different terms is that it seems at least possible, even likely, that our moral propensity would not in fact be equivalent to our moral disposition.<sup>7</sup> Rather, we might think that Kant's claim that they are equivalent is in fact a significant result of the radical evil thesis – our evil propensity corrupts us at the very root of our moral character. In any case, I shall follow Kant here in using the terms interchangeably.

Kant's radical evil thesis is the claim that there is a universal moral propensity to evil amongst the human species.<sup>8</sup> This means that "every human being, even the best", (6:32) has freely chosen to have an evil moral orientation as the subjective ground of their power of choice. Kant is using the term 'evil' in a broad sense to cover the entire range of human immorality, from telling a white lie to perpetrating genocide. As such, we

---

<sup>6</sup> This equation of *Gesinnung* and moral *Hang* has also been noted in Allison, *Kant's Theory of Freedom* 153, and Matthew Caswell, "Kant's Conception of the Highest Good, the *Gesinnung*, and the Theory of Radical Evil," *Kant-Studien* 97 (2006): 199.

<sup>7</sup> Indeed Bernstein argues that Kant owes us an explanation as to why a disposition can be either good or evil, but a propensity can only be evil - see Richard J Bernstein, *Radical Evil: A Philosophical Interrogation* (Cambridge: Blackwell Publishers, 2002) 26. Kant's radical evil thesis gives us the answer – in fact we all have (at least to start with) an evil disposition.

<sup>8</sup> Except perhaps Jesus who is, in any case, hardly the norm. (6:80)

might more accurately capture Kant's claim by talking of a universal human propensity to 'wrongdoing', where this is to be understood to include everything from doing the right thing for the wrong reason, to deliberately doing the wrong thing. However, in order to avoid confusion, I shall continue to use the term 'evil'.

Kant claims that the universality of the propensity to evil can be made compatible with the contingent nature of human freedom as it is a "*natural propensity*" which is "entwined in humanity itself, and, as it were, *rooted* in it". (6:32) Hence Kant is using the term 'radical' in the sense of 'rooted-in', and not in the sense of 'extreme'.<sup>9</sup> All human beings are radically evil for Kant because the propensity towards evil is so very deeply ingrained in human nature that it corrupts our power of choice at its very root. For Kant the "*radical innate evil* in human nature (not any the less brought upon us by ourselves)" (6:32) comprises our "*innate guilt*" and is "the foul stain of our species". (6:38) Kant claims that our propensity to evil comes in three different forms, which differ in grade but not in type, as each form is but a different manifestation of the same evil moral orientation, whereby sensuous incentives are made the condition of moral ones.

The first grade refers to the *frailty* of human nature when it comes to actually living up to our moral values. Even when we have recognised ideally what we ought to do, when it comes to implementing this in practice, especially when it is not to our advantage, we often find our moral commitments too frail to trump other interests. We are also frail when, without revoking our commitment to certain moral duties, such as to not steal, we make an exception for ourselves, while still maintaining that such duties are 'in general' valid. This often mires us in a pool of self-deception, whereby we rationalise away our frailty by attempting to justify actions which we know, at some level, to be unjustifiable.<sup>10</sup>

The second grade refers to the *impurity* of the human heart in its tendency to mix pure and impure incentives. Not only are our true motives for acting often (or always)

---

<sup>9</sup> Although there is at least one instance where Kant does use the term in the sense of 'extreme' – see Immanuel Kant, *Lectures on Ethics*, ed. Peter Heath and J B Schneewind, trans. Peter Heath (Cambridge: Cambridge University Press, 1997) 27:317. In this text the radically evil person is in fact equivalent to what Kant later refers to in the *Religion* as a "diabolical being". (6:35)

<sup>10</sup> Allison in particular has emphasised the importance of self-deception in Kant's account of evil – see Henry E Allison, "Reflections on the Banality of (Radical) Evil," in *Rethinking Evil*, ed. Maria Pia Lara (Berkeley: University of California Press, 2001), 90. However, it is important to note that not *all* evil is to be explained in terms of self-deception, a point I stress in a discussion of Eichmann – see Paul Formosa, "Moral Responsibility for Banal Evil," *Journal of Social Philosophy* 37, no. 4 (2006): 513.

opaque, even to ourselves, but we find it difficult, or even impossible, to ever act solely for the sake of the moral law. Kant gives the following example:

how would a man ascertain whether his joy at the rescue of an unfortunate family stems from sympathetic, pathological fellow-felling, or from the pleasure of the fulfilment of his duty, or whether, in his action, the love of honour or advantage did not obscurely play a part?<sup>11</sup>

We might give to charity, for example, not only because we think it is our duty, but at least in part because this makes us feel good about ourselves, or because we know it to be a tax-deduction, etc. Our best acts are often impure in that they are done, less from duty, than, for example, from a competitive desire to increase our social status in the eyes of others, or from an envious desire to catch up to (or bring down) those we perceive as 'above us'.<sup>12</sup> We rarely act, if ever, purely for the right reasons.

The third grade refers to the *depravity*, *corruption*, and *perversity* of the human heart,<sup>13</sup> as demonstrated by the human ability to intentionally "choose maxims that subordinate the incentives of the moral law to others (not moral ones)". (6:30) We are often not only too frail to follow the moral law, and when we do follow it, we do so only (or often) impurely, but we also sometimes perversely pursue courses of action that we know to be wrong. There are all sorts of reasons why we might do this.<sup>14</sup> We might do so, for example, because we find it in our self-interest, or because we harbour certain racist beliefs that incline us to evil, or even because we gain a sadistic pleasure from inflicting destruction on others.<sup>15</sup> Humans are the sorts of beings who are more than capable of intentionally and knowingly trampling all over their most sacred values, or even denying them, in order to perversely pursue ends in violation of the moral law. We

---

<sup>11</sup> From Kant's lectures on the *Metaphysics of Morals*, as recorded by Vigilantius, in Immanuel Kant, *Lectures on Ethics* 27:624. See also Immanuel Kant, "Groundwork of the Metaphysics of Morals," in *Practical Philosophy*, ed. Mary J Gregor (Cambridge: Cambridge University Press, 1996), 4:407.

<sup>12</sup> See, for example, Immanuel Kant, "Idea for a Universal History from a Cosmopolitan Point of View," in *On History*, ed. Lewis White Beck (Indianapolis: Bobbs-Merrill, 1963), 15-16, Kant, *Lectures on Ethics* 27:693.

<sup>13</sup> Kant uses all three terms as roughly equivalent. However, in order to be clear, I shall henceforth use only the term 'perversity' to refer to this third grade.

<sup>14</sup> I examine this question in detail in Paul Formosa, "Understanding Evil Acts," *Human Studies* (2007): forthcoming.

<sup>15</sup> See in particular Kant's account of the vices of hatred for human beings in Immanuel Kant, "The Metaphysics of Morals," in *Practical Philosophy*, ed. Mary J Gregor (Cambridge: Cambridge University Press, 1996), 4:458-62. Kant's account of passion (as opposed to affect) is also relevant here - see Immanuel Kant, *Anthropology from a Pragmatic Point of View*, ed. and trans. Robert B Loudon (New York: Cambridge University Press, 2006) 7:265-76.

are frail when we have to deceive ourselves about the evilness of our own acts, but we are perverse when such deception is not even necessary.

### 3. Kant's argument for radical evil

There has been much controversy in the secondary literature over what exactly Kant's argument for radical evil is, or whether he even can or does offer any such argument. In order to resolve these difficulties a close reading of Kant's text is required. By the third paragraph of Part One, Kant has already given the form of his argument:

In order, then, to call a human being evil, it must be possible to infer a priori from a number of consciously evil actions, or even from a single one, an underlying evil maxim, and, from this, the presence in the subject of a common ground, itself a maxim, of all particular morally evil maxims. (6:20)

Hence Kant's argument must be an a priori one.<sup>16</sup> Elsewhere Kant notes that the claim that humanity is evil by nature cannot "be inferred from the concept of his species (i.e. from the concept of a human being in general, for then the quality would be necessary)". (6:32) Hence the argument cannot be analytic. Therefore Kant's radical evil thesis can only be an a priori synthetic claim. As such, Kant's argumentative strategy is clear: he must proceed from the possibility of an evil maxim to the preconditions of such a maxim, namely to an evil propensity.

However, things might not be so simple, for in section III of Part One, Kant writes:

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. (6:32-3)

But this somewhat odd claim fits in neither with what Kant says elsewhere in the *Religion* nor with his critical philosophy more generally. An inductive generalisation cannot spare us the need for a deduction of a synthetic a priori claim.<sup>17</sup> Kant then goes on to write, in a footnote, also in section III of Part One:

---

<sup>16</sup> Wood simply dismisses Kant's claim here as "wildly implausible", because "how could it be possible to infer a priori from a *single* action an underlying ground of all morally evil maxims (not only for the agent but for the entire human species?)" – Allen Wood, *Kant's Ethical Thought* (Cambridge: Cambridge University Press, 1999) 286. But Kant argues only that we can infer from a single evil action an evil disposition, not in the entire species, but only in that individual. The universality claim requires a separate argument.

<sup>17</sup> Michalson notes that Kant's "appeal to experience" here "cannot possibly support the argumentative weight Kant seems to be placing on it", a point echoed by Loudon – see Gordon Michalson, *Fallen Freedom: Kant on*

the appropriate proof of this sentence of condemnation [of humanity's radical evil] by reason sitting in moral judgment is contained not in this section, but in the previous one. This section contains only the corroboration of the judgment through experience – though experience can never expose the root of evil in the supreme maxim of a free power of choice in relation to the law, for, as an *intelligible* deed, the maxim precedes all experience. – From this...we can also see why the principle of the exclusion of a mean between good and evil must be the basis of the intellectual judgment of humankind. (6:39)

From this the following is clear: 1) that Kant recognises that he needs to give a formal proof of his radical evil thesis, as experience can only corroborate such a proof and not spare us the need for it; 2) that Kant thinks that he has indeed given such a proof “by reason” in “the previous” section, i.e. in section II, and so *before* the above quoted claim that experience spares us the need for a proof; and 3) that Kant thinks his proof rests in part on the ‘rigorist’ view that there is no mean between an evil and a good propensity. Kant is again, however, misleading, for section II does not (at least seem to) contain the promised proof, but merely the formulation of the three grades of the evil propensity.

Does Kant deliver on the promised proof elsewhere in the text? He does not, at least explicitly, seem to do so. Kant does give parts of the required proof, but fails to explicitly bring the matter to any definite conclusion. Indeed Kant often goes out of his way to discuss humanity's evil propensity only as an unproven possibility, often prefacing his remarks with a hypothetical ‘if’ clause,<sup>18</sup> although it is also clear that he thinks our evil propensity is not only conceptually possible, but actually describes our human condition.

In any case, it is not too difficult to tie the threads together and construct an argument from Kant's text. There are four key elements to such an argument. First, there is Kant's so-called ‘rigorism’, which is the position that a human being's moral propensity can be exactly either good or evil. This claim is easy enough for Kant to defend, once it is understood that a propensity is a single maxim which, like all other maxims in which the moral incentive is present, is either good or evil but not both, depending on which incentive is incorporated as the supreme condition of that maxim. In

---

*Radical Evil and Moral Regeneration* (New York: Cambridge University Press, 1990) 46, and Robert B Loudon, *Kant's Impure Ethics: From Rational Beings to Human Beings* (Oxford: Oxford University Press, 2000) 221.

<sup>18</sup> To give but one example of many, consider in section IV, *after* the alleged proofs have been given, that Kant still says that “in order...to explain this propensity, *if there is one*”, (6:41, italics mine) thereby retaining only a hypothetical formulation of the thesis.



other words, Kant's rigorist view is simply that no single maxim can be *both* good and evil at the same time, and not that no single person can will both good and evil maxims. Second, there is Kant's so-called 'Incorporation Thesis', (6:44)<sup>19</sup> which is the view that the: "freedom of the power of choice has the characteristic...that it cannot be determined to action through any incentive *except so far as the human being has incorporated it into his maxim* (has made it into a universal rule for himself, according to which he wills to conduct himself)". (6:23-4) This claim is also easy for Kant to defend as it follows from the freedom of our power of choice that our free decisions can be determined by an incentive only to the extent that one chooses to incorporate that incentive into a maxim. In other words, from a practical point of view, when we represent the actions of agents as free, we do not think of incentives as *causing* the agent to act as they do, but rather we think of the agent as choosing to act on that incentive by freely *incorporating* it into their maxim. Third, in order to account for our ability to formulate evil maxims, by subordinating the incentive of the moral law, Kant needs to show that an evil propensity must be presupposed (6:20)<sup>20</sup> - call this the 'evil disposition derivation'. Fourth, Kant needs to argue that an evil moral propensity holds universally across the species - call this the 'universality claim'. It is only the last two parts of this argument that require further investigation.

#### 4. The evil disposition derivation

Kant's conception of a disposition has been seen by many to be an important addition or clarification to his moral theory.<sup>21</sup> It is through the concept of a disposition that we can, through tying the threads of one's individual decisions together by seeing them as an expression of one's underlying moral character, move beyond viewing agency simply in terms of single isolated acts of the will. As Allison explains:

the choices of rational agents, or in his terms, the maxims they adopt, must be conceived in relation to an underlying set of intentions, beliefs, interests and so on, which collectively

---

<sup>19</sup> See Allison, *Kant's Theory of Freedom* 5.

<sup>20</sup> This claim is also made in Henry E Allison, "On the Very Idea of a Propensity to Evil," *The Journal of Value Inquiry* 36 (2002): 341, Allison, "Reflections on the Banality of (Radical) Evil," 87, Sharon Anderson-Gold, *Unnecessary Evil: History and Moral Progress in the Philosophy of Immanuel Kant* (New York: State University of New York Press, 2001) 40, Emil Fackenheim, "Kant and Radical Evil," *University of Toronto Quarterly* XXIII (1954): 349.

<sup>21</sup> See John R Silber, "The Ethical Significance of Kant's Religion," in *Religion within the Limits of Reason Alone* (New York: Harper Torchbooks, 1960), cxiv.

constitute that agent's disposition or character. Otherwise these choices and maxims could neither be imputed nor explained.<sup>22</sup>

This sounds all very well and good, but in fact Allison is somewhat misleading here. This is because for Kant our disposition per se is *not* made up of a collective web of intentions, beliefs and interests. Such a gloss on Kant's concept of a disposition misses what is most distinctive, and also most counter-intuitive, about his account. For Kant our disposition is encapsulated in a single 'supreme maxim', for which we are responsible and yet which we innately acquire 'at birth'. (6:22, 25)

This raises an obvious problem – how can I be said to choose my character via some supreme maxim at birth? Allison notes that implicit in many of the “standard attacks on Kant's conception of freedom is the assumption that in choosing a *Gesinnung* one is in effect choosing a phenomenal self in toto, or at least a 'character' in the usual sense of a relatively fixed nature, construed roughly along the lines of the Aristotelian *hexis*”.<sup>23</sup> Clearly, in order for us to be completely responsible for our disposition or character, Kant cannot conceive of a disposition anywhere nearly as 'thick' as the Aristotelian *hexis*. But then the problem is that, if we conceive of a disposition too 'thinly,' as Kant seems to do in terms of a single supreme maxim, then it is no longer clear that such an account can do the job it needs to do – namely, account for the coherency and continuity of moral agency.

The way to avoid this dilemma is to see that the coherency and continuity that Kant's account of disposition provides is achieved, not solely in terms of a single supreme maxim, but in terms of the hierarchical understanding of maxims that such a disposition presupposes. Kant's conception of a disposition does not, then, refer to our thick psychological states, habitual inclinations and the like, but only to our fundamental moral orientation as specified in our most general and supreme maxim. And this supreme maxim in turn makes sense only against a background hierarchy of maxims, of various degrees of generality, which encapsulate the incorporated values, beliefs, interests and intentions that an agent has over a lifetime.

There is of course much debate about what, for Kant, does and does not count as a maxim. Some have argued that only general life principles count as maxims, others that everyday intentions count as maxims, while others have argued for a position

---

<sup>22</sup> Allison, *Kant's Theory of Freedom* 136.

<sup>23</sup> *Ibid.* 141.

somewhere in between.<sup>24</sup> But whatever level of generality we conceive maxims to have, Kant's account of a disposition clearly commits him to the view that there is a hierarchy of maxims. This is because Kant is committed to the view that there is at least a two-level hierarchy, in terms of the supreme maxim that is our disposition and the maxims that it grounds.<sup>25</sup> Whether or not these lower-order maxims are arranged flatly (as a 'life rules' reading might imply) or are themselves arranged in a complex tree hierarchy (as an 'everyday intentions' reading might imply), can be left open here – but in either case, hierarchy there must be, whether it be two-level or multi-level. As such, I shall discuss here only the multi-level case, whereby we have maxims of various levels of generality arranged in a tree hierarchy. I shall do so, not only because this is arguably the correct way to read Kant, but also because it is the harder of the two cases to handle, and thus if the evil disposition derivation can be defended for this case, then a similar argument will also hold for the two-level case.

But how do higher-order maxims relate to lower-order maxims? Consider the case of Matthew, who adopts the maxim to obtain shelter during the winter.<sup>26</sup> This maxim leads Matthew to adopt lower-order maxims to build a house, to hammer on the roof, to buy nails and so on. When someone sees Matthew buying nails we can imagine the following dialogue ensuing: 'why are you doing that?' – 'because I want to hammer on the roof' – 'why?' – 'because I want to build a house' – 'why?' – 'because I want to obtain shelter during the winter'. Of course none of this deliberation need explicitly pre-date the deed of buying nails, but rather, in buying the nails Matthew shows that he is implicitly committed to such maxims. As Caswell puts it: "my visible behavior, insofar as we consider it a free, purposive act, presupposes a system of policies that supply my acts

---

<sup>24</sup> Allison critically examines the views of Höffe and Bittner (only 'life rules' count as maxims) and O'Neill (only 'underlying intentions' count as maxims) and argues that "Kantian maxims come in various degrees of generality" and that "one might think of maxims...as arranged hierarchically, with the more general embedded in the more specific" - Ibid. 91-4. Caswell also argues that it seems reasonable to think some maxims will be momentous life decisions, others less so or not at all –see Caswell, "Kant's Conception of the Highest Good, the *Gesinnung*, and the Theory of Radical Evil," 193.

<sup>25</sup> Kant's talk of the "very highest maxim in us" (6:32) clearly implies a hierarchy account of maxims. Allison also argues that "this hierarchical view of maxims is presupposed by Kant's conception of *Gesinnung* and radical evil, which rest on the assumption of a fundamental maxim underlying the choice of more specific maxims" - Allison, *Kant's Theory of Freedom* 94.

<sup>26</sup> I take this example from Caswell, "Kant's Conception of the Highest Good, the *Gesinnung*, and the Theory of Radical Evil," 193.

their rational justification".<sup>27</sup> As such, a higher-order maxim can be said to 'ground' a lower order maxim in the sense that it provides a rational justification or explanation (i.e. a grounding) for that maxim. The maxim to buy nails, for example, only makes sense, and can only be considered a purposive act, against the background of a higher-order maxim to build a house, and such a maxim in turn only makes sense against the background of an even higher-order maxim to obtain the necessities of life and so on. Every maxim requires a grounding in some higher-order maxim, just to the extent that we consider people to be free moral agents, with underlying intentions, beliefs, interests and the like, engaging in purposive acts under the requirements of practical reasoning.<sup>28</sup>

However, it is simply a mistake to think that on this model one must first formulate their highest-order maxim, then the next one down and so on, to the lowest. Rather "the higher-order maxim is rationally presupposed by the lower, not temporally preceded".<sup>29</sup> Thus higher-order maxims do not undermine one's spontaneity by determining lower-order maxims. For example, Matthew might just as well decide to adopt a maxim to seek shelter with friends, rather than build a house, in order to achieve his end of obtaining shelter during the winter. But while higher-order maxims do not determine, they do constrain our choice of lower-order maxims. For example, deciding to daydream whenever it gets cold is not in any way a means to the end of obtaining shelter during winter, and so it cannot be rationally explained or justified in terms of that particular higher-order maxim.

With this conception of practical reasoning in place it is now possible to defend the coherency of Kant's sometimes baffling account of a moral propensity or disposition. For Kant a moral propensity is the "subjective determining ground of the power of choice *that precedes every deed*, and hence is itself not yet a deed". (6:31) Hence, for Kant this subjective determining ground must be "posited as the ground antecedent to every use of freedom given in experience (from the earliest youth as far back as birth) and is thus represented as present in the human being at the moment of birth – not that birth itself is its cause". (6:22) This, as Kant assures us, does not mean that this supreme maxim has "not been earned by the human being who harbours it, i.e. that he is not its author,

---

<sup>27</sup> Ibid.: 195.

<sup>28</sup> This claim requires two significant clauses: except in the special case of the dispositional maxim, which is the highest-order maxim, and all morally good maxims, which provide their own ground. I discuss these cases in detail later.

<sup>29</sup> Caswell, "Kant's Conception of the Highest Good, the *Gesinnung*, and the Theory of Radical Evil," 195-6.

but means rather that it has not been earned in time." (6:25) We might summarise Kant's view as follows: our moral propensity is the result of a timeless deed that precedes all other deeds, is present at the moment of birth and has already then been earned by the human being. Is this even a coherent position?<sup>30</sup>

If we read Kant as offering the view that at the very moment of birth (how long after conception?) we make a free, timeless noumenal choice to adopt an evil supreme maxim, then his position is metaphysically daunting and wildly implausible.<sup>31</sup> But there is no reason to read Kant in this overtly metaphysical way. Kant only claims that such a supreme choice must be "posited" and thus "represented" as being present at birth. It is not as if we adopt our supreme maxim first, at birth say, and then reason downward. Rather the reverse is the case. We begin to use our freedom by adopting some lower-level and unimportant maxim, such as to 'obtain shelter during the winter'. But *any* maxim already presupposes a complex hierarchy of maxims, in terms of which that maxim can be understood, which the agent may not, indeed is very likely not to be, explicitly aware of at the time.<sup>32</sup> In order to see an agent's actions as purposively guided by practical rationality, we need to 'posit' or 'represent' the agent as *already* committed to various higher-order maxims and ultimately a certain dispositional maxim. This dispositional maxim must be posited, in order to account for the coherency and continuity of agency, as already implicitly adopted by the agent before they begin adopting other maxims.

From this it is clear that Kant's position is not, or at least need not be read as, either paradoxical or incoherent. To avoid this charge it needs to be shown that a

---

<sup>30</sup> For reasons like this Bernstein sees Kant as at "war with himself" - Bernstein, *Radical Evil: A Philosophical Interrogation* 11, 33. Sussman sees Kant's radical evil thesis as a "mess of paradoxes". He notes: "It is an inextricable aspect of human nature; yet it is the product of our individual exercises of freedom. Radical evil seems to be a priori necessary to each and every member of the species; yet it is inessential to humanity" - David G Sussman, *The Idea of Humanity: Anthropology and Anthroponomy in Kant's Ethics* (New York: Routledge, 2001) 227. Kant himself raises the concern that there might "be a contradiction in the concept of a simple propensity to evil," and he admits that there are legitimate concerns that this propensity may not be "reconcilable with the concept of freedom". (6:31)

<sup>31</sup> As Sussman notes: "The idea of a noumenal choice, made outside of time, for all time, by a contentless noumenal agent seems completely unintelligible" - Sussman, *The Idea of Humanity: Anthropology and Anthroponomy in Kant's Ethics* 13.

<sup>32</sup> All that is required is that it "must be possible in subsequent reflection to discover and articulate...the maxims on which one acts" - Allison, *Kant's Theory of Freedom* 90. See also Allison, "On the Very Idea of a Propensity to Evil," 341.

disposition can be *both* freely chosen and innate. (6:22, 25) One way to do this is to apply Kant's distinction between a physical and a moral propensity to the term 'innate'. Thus something can be either physically innate (rooted in our unchosen nature) or morally innate (rooted in our freedom). A moral propensity is morally and not physically innate, because it alone amongst all our maxims must be presupposed, in order to account for the coherency and continuity of agency, as present from the very first use we make of our freedom until the last. A disposition, then, is innate to our freedom, and not to our physical birth. Thus we must interpret Kant's claim that we have our disposition 'at birth' to mean 'at the birth of our freedom', which Kant thinks occurs somewhere around our twentieth year.<sup>33</sup>

Similarly, the 'timelessness' of our disposition can equally be understood in a more common sense way. While of course an agent adopts different maxims at different times, such maxims are always adopted for practical reasons. Or at least, this is how we must represent it, from a practical point of view, if we are to fully impute an action to a person, which we do by thinking of them as moral agents with practical rationality acting under the demands of the categorical imperative. Hence by the 'timelessness' of our disposition, Kant means no more than that, as with any other maxim, a disposition is not to be understood in terms of causes and effects in time, but rather in terms of a process of practical rationality. As such, all maxims, not just dispositional maxims, are 'timeless' in this weak sense. This of course relies on Kant's account of freedom, which may have problems of its own, but the important point here is that Kant's conception of a disposition adds no new difficulties. Properly understood then, there seems to be nothing incoherent in the very idea of a disposition.

We are now in a position to present Kant's evil disposition derivation. The argument is, given the above discussion, simple and straightforward. Take the case of any human agent with any evil maxim. That maxim, insofar as it is evil, must be one that has incorporated self-love as its supreme condition and thus subordinated the moral

---

<sup>33</sup> This fits in with what Kant says elsewhere - see Immanuel Kant, "Conjectural Beginning of Human History," in *On History*, ed. Lewis White Beck (Indianapolis: Bobbs-Merrill, 1963), 60, Immanuel Kant, "Lectures on the Philosophical Doctrine of Religion," in *Religion and Rational Theology*, ed. Allen Wood and George Di Giovanni (Cambridge: Cambridge University Press, 1996), 28:1078. In both these texts Kant makes it clear that practical freedom emerges, and equally ethical obligation commences, only when we reach maturity. If our moral propensity is something for which we are responsible, we cannot possibly acquire it before we acquire our freedom, which in turn is acquired at maturity and not birth.

incentive. But this is only possible on the presumption that the agent has an evil disposition. This is so because an agent's disposition, or their supreme or most general maxim, is, given Kant's rigorism, exactly either good or evil. If the disposition is good then an evil maxim is impossible. This is because a good disposition is one that makes its supreme maxim the rule that morality is *always* to be unconditionally followed. Such a supreme maxim cannot therefore ground *any* maxim that is rationally incompatible with this disposition. But *any* maxim that is evil makes morality conditional and is therefore incompatible with a good supreme maxim. Therefore only an evil disposition can rationally ground an evil maxim.<sup>34</sup> This argument thus shows that from any particular evil maxim it can be shown that an evil disposition must be presupposed as providing the rational grounds for that maxim.

However, even though an agent with a good disposition cannot will an evil maxim (without repudiating that disposition), an agent with an evil disposition can will *both* good and evil maxims. Why this asymmetry? Because only evil maxims need to be necessarily traced up the maxim tree, so to speak, whereas good maxims are adopted, insofar as they are good, for their own sake, that is, for their mere universal form. There is no *extra* reason for adopting a moral maxim, beyond the maxim's form, whereas there is always an *extra* reason for adopting an evil maxim. That is, unlike good maxims, evil maxims are adopted for the sake of an end which must be understood against the background of an agent's beliefs, desires, passions, intentions etc, as incorporated into that agent's maxim tree. In other words, good maxims are anomalies of sorts, in that they are adopted not for the sake of obtaining some end peculiar to a particular agent, but from the respect that the mere consciousness of the maxim being in conformity with the moral law engenders.<sup>35</sup>

Thus for Kant even a "human being of good morals (*bene moratus*)," whose actions may always accord with the letter of the law, but who does "not always have,

---

<sup>34</sup> A similar argument is found in Caswell, "Kant's Conception of the Highest Good, the *Gesinnung*, and the Theory of Radical Evil," 206.

<sup>35</sup> The anonymous reviewer of this paper raised the following objection: either all maxims require a ground, in which case good maxims do require an 'extra reason', or not all maxims require a ground, in which case the evil disposition derivation fails – one can't have it both ways. Barring complications with the dispositional maxim, which I discuss later, the way out of this difficulty is to see that good maxims alone provide their *own* ground, and this is why we act autonomously only when we act morally. Conversely, we act heteronomously when we act immorally, for in such cases the reason why we act as we do is ultimately dependent on something other than our own pure practical reason.

perhaps never have, the law as their sole and supreme incentive" (6:30) has an evil disposition. Equally a villain who never acts in accord with even the letter of the law has an evil disposition. Thus any person who does not *always* make the moral law their sole and supreme incentive for adopting a maxim has an evil disposition. In contrast, Kant calls a "morally good human being (*moraliter bonus*)" one who "*always*" has the "law as their sole and supreme incentive," (6:30) and so has a good disposition.

But a morally good human being would not be equivalent, as Caswell thinks, to the "finite *holy* beings (who could never be tempted to violate duty)" that Kant writes of in the *Doctrine of Virtue*.<sup>36</sup> For Kant "human morality in its highest stage can still be nothing more than virtue, even if it be entirely pure". This is because even those humans who are "entirely pure" (i.e. who have a good disposition), must remain in a state requiring also the "*autocracy* of practical reason", not just the "autonomy of practical reason", for a good disposition ensures only the "*capacity* to master one's inclinations when they rebel against the law".<sup>37</sup> This capacity is based on the fact that rebelling inclinations, i.e. temptations, must be incorporated into a maxim, something a human with a good disposition could never do (without repudiating that disposition). Thus humans, even those (if there are any) with good dispositions, would remain, unlike a finite holy being, subject to temptations,<sup>38</sup> even if they never gave into those temptations.

However, there is one important objection that still needs to be addressed. If every maxim makes sense only within a broader background of higher-order maxims, then it would seem to follow that we would require an even higher-order level maxim to ground our supreme maxim, and an even higher-order maxim to ground our maxim that grounds our supreme maxim and so on. This would lead to an infinite regress.<sup>39</sup> But this objection does not hold. This is because maxims cannot be of an ever higher and higher level of generality. All maxims involve an ordering of incentives. (6:36) Thus the most general maxim possible is one that does not order particular incentives on a particular occasion, but that orders the different *types* of incentives per se, in a situation-independent way. This most general maxim would define one's moral orientation or

---

<sup>36</sup> See Caswell, "Kant's Conception of the Highest Good, the *Gesinnung*, and the Theory of Radical Evil," 202.

<sup>37</sup> All quotes here from Kant, "The Metaphysics of Morals," 6:383.

<sup>38</sup> Which Caswell denies - see Caswell, "Kant's Conception of the Highest Good, the *Gesinnung*, and the Theory of Radical Evil," 202.

<sup>39</sup> See Michalson, *Fallen Freedom: Kant on Radical Evil and Moral Regeneration* 59.



disposition by making either sensuously derived incentives, or purely moral incentives, the condition of the other. Therefore it is impossible to conceive of a maxim that could ground our most general maxim, for no maxim could be more general, and thus infinite regress worries are unfounded.

But this leads to another problem. Given that each and every maxim is spontaneously adopted as an act of freedom, it follows that why an agent adopts one maxim and not another is always in some sense inscrutable.<sup>40</sup> Nonetheless, why an agent adopts a particular maxim can still be made sense of and understood against a background of higher-order maxims which encapsulate that agent's values, beliefs, intentions and the like. But if we understand maxims in this way, then eventually we will hit rock-bottom, and we have hit rock-bottom when we reach our dispositional maxim. There is nothing above such a maxim in terms of which it can be justified. Thus our moral 'fall' is inscrutable in a deep sense, though not due to the mysterious nature of evil, but because our supreme maxim is our explanatory rock-bottom. As such, a good propensity would be equally inscrutable.

## 5. The universality claim

In the previous section I defended Kant's evil disposition derivation. From a single evil maxim we can infer an underlying evil disposition. The next and final stage of the proof for Kant's radical evil thesis is to progress from this claim to the one that all humans, even the best, have an evil disposition. It is here, though, that Kant's argument fails. But not only that, I shall also show that Kant was fully aware of the provisional nature of his thesis.

Kant's argument for the universality claim is as follows:

That by the 'human being' of whom we say that he is good or evil by nature we are entitled to understand not individuals (for otherwise one human being could be assumed to be good, and another evil, by nature) but the whole species, this can only be demonstrated later on, if it transpires from anthropological research that the grounds that justify us in attributing one of

---

<sup>40</sup> Kant argues that "we can explain what happens only by deriving it from a cause in accordance with laws of nature, and in so doing we would not be thinking of choice as free" - Kant, "The Metaphysics of Morals," 6:380. Hence all free choices are inscrutable in the sense that they cannot be explained in terms of prior causes. Of course, when we act freely we act for a *reason*, but why it is we act for *that* reason, is in some sense inscrutable, just insofar as we are free.

these two characters to a human being as innate are of such a nature that there is no cause for exempting anyone from it, and that the character therefore applies to the species. (6:25)

But no amount of anthropological evidence can give us grounds for proving that no single human being has ever started out with a good disposition. There are a number of reasons why this is so. First, we can have no direct experience of dispositions, so that if an agent performs no evil acts we cannot be certain that they have an evil disposition – indeed, the evidence in such a case would seem to suggest rather that they had a good disposition. Second, Kant, as the author of the first *Critique*, is well aware that an inductive generalisation from anthropological evidence cannot support a claim to universality. For that a transcendental deduction is required. As such, anthropological research can at best establish only that evil has *so far* been *widespread*, and not that it is universal.<sup>41</sup> Third, the anthropological evidence itself, insofar as it can provisionally support any position, supports not Kant's position, but rather the position that an evil disposition is widespread, such that most, but not all, humans are radically evil. There is evidence to suggest that there have been at least some moral saints who were pure in disposition.

However, later in the text Kant argues that:

A member of the English Parliament exclaimed in the heat of debate: 'Every man has his price, for which he sells himself.' *If this is true (and everyone can decide by himself)*, if nowhere is a virtue which *no level of temptation can overthrow*, if whether the good or evil spirits wins us over only depends on which bids the most and affords the promptest pay-off, then, what the Apostle says *might indeed hold true* of human beings universally, 'There is no distinction here, they are all under sin – there is no one righteous (in the spirit of the law), no, not one'. (6:38-9 – italics mine)

A person with a good disposition is one who is righteous in "the spirit of the law" and possesses a virtue which "no level of temptation can overthrow". Such a person has no price. No level of temptation will ever be sufficient to prompt them to adopt an evil maxim. Has there ever been such a person? Kant's rhetoric here is no doubt of immense power but, as a proof, it fails. For how can one prove that there can be no such person?

Indeed, we have reason to think, rare as they might be, that there have been a few examples of agents with good dispositions. An agent with a good disposition is one

---

<sup>41</sup> As Allison notes: "the *most* that this [anthropological] evidence can show is that evil is widespread, not that there is a universal propensity to it" - Allison, *Kant's Theory of Freedom* 154. Grimm makes a similar point - Stephen R Grimm, "Kant's Argument for Radical Evil," *European Journal of Philosophy* 10, no. 2 (2002): 165.

who is never (and will never be) frail, impure or perverse. While perhaps many humans have managed to satisfy the third condition, by never deliberately adopting perverse maxims, the frailty and impurity conditions are far harder to meet. In order to avoid the charge of possessing an evil disposition, an agent would have to at least,<sup>42</sup> beyond never performing wrongful acts, always and only have adopted morally good maxims and actually acted upon them (the frailty condition), and have always adopted those maxims for duty's sake (the impurity condition).<sup>43</sup>

But who has never been impure? Who has never told a white lie in order to make things easier for themselves, or to avoid hurting someone's feelings? Who has never helped a friend from (rather than just with) sympathy or love, and thus not from duty, where to help one's friend was morally obligatory? And who can be sure that they never have? The frailty requirement is even more difficult to meet. Kant, in order to illustrate this point, draws on the macabre example of Phalaris' bull. (6:50)<sup>44</sup> Phalaris' bull refers to the tyrant of Agrigent's preferred method of torture, namely of placing his enemies inside a hollow cast iron ore bull, which was heated red hot. The test Kant has in mind is clear: if Phalaris gave you the horrible choice between committing perjury and being tortured in his bull, which would you choose? To choose the former is for Kant, though perfectly understandable, an example of human frailty and evidence of an evil disposition. But would *any* human not be frail in such a circumstance?

History has furnished us with a few immensely precious examples of those who have in good faith chosen to die rather than do wrong. How many of us could say with Robert Bolt's Thomas More: "I do none harm, I say none harm, I think none harm. And if this be not enough to keep a man alive, in good faith I long not to live"<sup>45</sup> Or say with Socrates: "I will not yield to any man contrary to what is right...even if I should die at once for not yielding"<sup>46</sup> Rare as such moral saints may be, that such humans cannot or

---

<sup>42</sup> Note the 'at least', for even this would be no absolute guarantee of a good disposition.

<sup>43</sup> However, what exactly is involved in 'acting from duty' is a matter of debate. Depending on what one took this to mean, such a condition might be more or less taxing. See Phillips Stratton-Lake, "Moral Motivation in Kant," in *A Companion to Kant*, ed. Graham Bird (Oxford: Blackwell, 2006).

<sup>44</sup> "(though Phalaris himself should command you to be false and, having brought up his bull, should dictate perjuries). Everybody must admit that *he does not* know whether, were such a situation to arise, he would not waver in his resolve." (6:50)

<sup>45</sup> Robert Bolt, *A Man for All Seasons* (London: Heinemann, 1960) 97.

<sup>46</sup> Plato, *Apology*, trans. G. M. A. Grube (Indianapolis: Hackett, 2002) 32a.

have never existed, seems untrue.<sup>47</sup> And if our moral saint also has (and has always had) purity, that is, possesses an ethical frame of mind whereby she always acts from duty, then such a person is not radically evil. The possibility of such a case, rare though it may be, is enough to undermine the universality claim.<sup>48</sup>

But Kant himself, or so I suggest, was fully aware of this, for he only states that “*if this is true*”, which “everyone can decide by himself”, (6:38) and not something along the lines of, ‘as I have already proven’. Thus it is not without reason that Kant often discusses humanity’s radical evil only in hypothetical terms. Even so, Kant’s thesis remains eminently plausible and there is more than enough evidence to suggest that an evil disposition is certainly widespread, and *perhaps* universally so, amongst humanity. Not only do the very vast majority of us have our price, but that price is for most of us embarrassingly low.

It is surely significant that in all other references to human nature, outside of the *Religion*, Kant talks only of impurity and frailty, and never of perversity.<sup>49</sup> While many (or even all) of us have at times a propensity to be frail or impure, and this is *alone* enough to indicate an evil disposition, arguably many of us never have any propensity whatsoever to be perverse, in the sense of deliberately adopting an immoral maxim in the light of the fact that it is immoral. So why think humans universally have a propensity to, not just frailty and impurity, but also perversity? An evil disposition is, in and of itself, sufficient to indicate, not that the bearer of that disposition necessarily has a propensity to all three grades of evil, but only that they might. As such, while most (if not all) humans surely have a propensity to be frail and impure, and this is enough to

---

<sup>47</sup> As Marcia Baron argues, the Kantian moral saint is not one whose life is solely “dominated by a commitment to improving the welfare of others”. The Kantian moral saint pursues, amongst other things, the end of their own self-perfection, as well as the happiness of others, and cultivates the sort of empirical character that provides them with the sensitivity to others, and the emotional resources required, to successfully undertake such a task – see Marcia Baron, “Moral Paragons and the Metaphysics of Morals,” in *A Companion to Kant*, ed. Graham Bird (Oxford: Blackwell, 2006).

<sup>48</sup> The anonymous reviewer of this article argues that while we cannot absolutely *rule out* such cases, we cannot absolutely *rule them in* either. However, the force of this argument can be lessened by focusing on cases like those of Thomas More or Socrates. If a person does not do wrong in a situation where their very life is at stake, then we have good reasons to think (though not with absolute certainty) that they must have a good disposition, for if their evil disposition were to shine through at any time, it would surely shine through then.

<sup>49</sup> See, for example, both the 1784-5 and the 1793 lectures in Kant, *Lectures on Ethics* 27:293, 27:350, 27:570-2 and Kant, “Groundwork of the Metaphysics of Morals,” 4:406.

indicate that they are radically evil, they need not necessarily also have a propensity to be perverse (although some, or even many, will).

In any case, the failure of the general universality claim, even granted the success of the claim that an evil disposition is widespread, is a considerable blow to Kant's thesis. This is because without the universality claim Kant cannot maintain that the character of the human *species* is evil. Thus if most, but not all, humans have an evil disposition, then the human species has no uniform moral character. As such, Kant can no longer maintain that radical evil is "the foul stain of our species," (6:38) even if it be a stain which at least the vast majority of us humans muddy ourselves unnecessarily with. But this has a significant upshot – it makes Kant's (revised) radical evil thesis more plausible and avoids certain difficulties. The main difficulty that it avoids is Kant's hopelessly cryptic attempt to justify the position that the universality claim is compatible with his account of freedom. Kant argues that:

because of freedom, such maxims must be viewed as accidental, a circumstance that would not square with the universality of evil at issue unless their supreme subjective ground were not in all cases somehow entwined with humanity itself and, as it were, rooted in it. (6:32)

However, if the choice of a disposition is genuinely free, it is not clear how its being, somehow, "entwined with humanity", whatever that means, can ensure that we all choose in the same way without undermining the freedom of that choice. If instead we admit that the universality claim does not necessarily hold, and it is therefore possible that not everyone starts out with an evil disposition (although most do), then there is no difficulty in asserting that while the cards are stacked in favour of one choice (more on this below), as a free choice, it remains possible to choose otherwise (and indeed perhaps some saintly humans have done just this).

While Kant's actual argument for the universality claim fails, as Kant himself arguably realised, there are a number of other possible defences that might be mounted. Caswell, for example, argues that an "evil orientation [i.e. propensity] can be inferred to belong to any will for which morality is a struggle against temptation".<sup>50</sup> However, even if we grant the premise that universally all humans, even the very best, are subject to temptations (a seemingly undeniable claim), the universality claim does not follow. This is because, as I argued above, the fact that a human being is subject to temptations is

---

<sup>50</sup> Caswell, "The Value of Humanity and Kant's Conception of Evil," 637. Allison also makes the claim that for a being with a good propensity "there could be no possibility of temptation" - Henry E Allison, "Ethics, Evil and Anthropology in Kant: Remarks on Allen Wood's Kant's Ethical Thought," *Ethics* 111, no. 3 (2001): 609.

not in itself indicative of an evil disposition. What makes the difference between a good and an evil disposition has nothing to do with whether or not the agent has tempting inclinations (that is, inclinations contrary to the law), but whether or not that agent incorporates those temptations unconditionally into their maxims. A good disposition cannot make one immune to the natural pull that the inclinations exert on us as finite sensible beings. Morality might indeed remain a virtuous struggle for even those who have a pure disposition but, unlike the rest of us, it is a struggle in which such agents can have complete confidence that they can always win, come what may.

Allison offers the similar argument that a good disposition is beyond human nature since no human can "spontaneously" "subordinate self-love to duty...without, as it were, giving self-love a hearing". For Allison this impossibility is due neither to our finitude, nor to our sensuous nature per se, but is rather a consequence of the "status of happiness" for Kant as a "necessary end for human beings".<sup>51</sup> Thus we cannot *spontaneously* choose duty over inclination, although with a struggle we can do so reluctantly, for we must first consider the claims of happiness, which is thus a source of resistance to morality whenever morality infringes upon it.

There are at least two ways to challenge this argument. First, one might just deny that there has never been a human moral saint whom, on at least some occasions, spontaneously chose duty over inclination, even at the expense of their happiness. But this, as a simple denial of Allison's assertion, gets us only so far. Second, the issue of spontaneity is not as decisive as Allison needs it to be. Does the mere fact that I do not adopt a good maxim in a spontaneous fashion, without first giving happiness a hearing, necessarily indicate an evil disposition? I would suggest not. What matters morally, and what alone is truly indicative of my moral disposition, is the ordering of the incentives which I actually incorporate into my maxims (and this, of course, our power of choice *always* does spontaneously), not the mere fact that when deliberating I considered choosing otherwise. As Kant writes: "those inclinations [which oppose the law] only make more difficult the *execution* of the good maxims opposing them; whereas genuine evil consists in our *will* not to resist the inclinations when they invite transgression, and this disposition is the really true enemy". (6:59) It is neither considering the natural end of happiness, insofar as it opposes the law, nor the non-spontaneous failure to choose duty

---

<sup>51</sup> See Allison, "On the Very Idea of a Propensity to Evil," 342.

above all else, which is indicative of an evil disposition, but only the *will* not to resist those inclinations by actually adopting evil maxims.

In support of his argument Allison refers to 5:83-6 of the second *Critique*. Here Kant contrasts a disposition of *respect* for the law with a disposition of (practical) *love* for the law. But *both* of these dispositions are good dispositions, as both make the moral law supremely unconditional. Kant calls the latter "an ideal of holiness" which is "not attainable by any creature but is yet the archetype which we should strive to approach and resemble in an uninterrupted but endless progress".<sup>52</sup> Holiness in disposition can be attained by no creature, for as creatures we:

can never be altogether free from desires and inclinations which, because they rest on physical causes, do not of themselves accord with the moral law...and consequently, with reference to these desires, it is always necessary for him to base the disposition of his maxims on moral necessitation, not on ready fidelity but on respect, which *demand*s compliance with the law even though this is done reluctantly.<sup>53</sup>

If "mere love for the law" were to be an obtainable disposition for sensuous creatures like ourselves, then morality would "cease to be a *command*, and morality, having passed subjectively into holiness, would cease to be *virtue*".<sup>54</sup> But a disposition of respect (not love) for the law is still a good disposition, as it never allows self-love to be the condition of morality, even if this can only be done reluctantly and non-spontaneously. Thus good dispositions can come in either virtuous or holy forms. Humans, being finite sensible creatures, can only achieve the virtuous form of a good disposition, as we remain always subject to temptations and inclinations that potentially oppose the law, but which can be and always are (thanks to one's disposition) mastered. The mistake common to both Caswell's and Allison's defence of the universality claim is to equate a good disposition with the ideal of holiness (even if only in its finite form) and then show that holiness (rather than a good disposition) is universally unattainable by humanity.

A different line of reasoning in support of the universality claim, which I shall call the 'development argument', has been forwarded independently by Sussman and Grimm.

---

<sup>52</sup> Immanuel Kant, "Critique of Practical Reason," in *Practical Philosophy*, ed. Mary J Gregor (Cambridge: Cambridge University Press, 1996), 5:83.

<sup>53</sup> *Ibid.*, 5:83-4.

<sup>54</sup> *Ibid.*, 5:84.

Much of this argument is derivable from Kant's reflections on pedagogy and the role of education in cultivating morality.<sup>55</sup> Sussman argues that:

the radical evil in human nature is a consequence of the fact that our empirical development in time does not correspond to the proper conceptual order [i.e. from noumenal to phenomenal] immanent in the idea of a person...Humans possess radical evil as a result of the fact that as creatures who must grow into their autonomy, we always start out from some condition in which the passions dominate; a condition which, if viewed narrowly, could only be deemed evil.<sup>56</sup>

Similarly, Grimm argues that, as we necessarily start from a state where "the animal element of our nature basically determines our behaviour," when we come to adopt our supreme maxim, "we choose to reject the law in favour of the needs and inclinations which were a natural part of our identity before the coming of the law".<sup>57</sup>

This development argument draws attention to the fact that humans universally start out as immature and unfree beings who begin life as the playthings of their inclinations. The problem is, as Kant puts it:

nature within man tries to lead him from culture to morality and not (as reason prescribes) from morality and its law [to culture]...And this course inevitably perverts his tendency and turns it against its end.<sup>58</sup>

Our radical evil is not a consequence, as Michalson claims, of the mere fact that we have "bodies",<sup>59</sup> but rather of the fact that we adopt a maxim to give our sensuous interests *priority* over moral ones. Kant's point is not that we should repress our sensuous natures (which Kant argues would be blameworthy),<sup>60</sup> but that we must develop from a state of

---

<sup>55</sup> See also Loudon, *Kant's Impure Ethics: From Rational Beings to Human Beings* 33-44, G. Felicitas Munzel, *Kant's Conception of Moral Character* (Chicago: The University of Chicago Press, 1999) 135, 308. Of Kant's texts see, for example, the following: Kant, "Conjectural Beginning of Human History," 60, Kant, "Critique of Practical Reason," 5:146-7, Kant, "Lectures on the Philosophical Doctrine of Religion," 28:1077-8.

<sup>56</sup> Sussman, *The Idea of Humanity: Anthropology and Anthroponomy in Kant's Ethics* 18.

<sup>57</sup> Grimm, "Kant's Argument for Radical Evil," 164, 71. Grimm's account seems to suggest that our first act of freedom is to *explicitly* adopt an ethical orientation. I have argued above that this is a misleading way to think about how a supreme maxim is adopted.

<sup>58</sup> Kant, *Anthropology from a Pragmatic Point of View* 7:327-8.

<sup>59</sup> See Michalson, *Fallen Freedom: Kant on Radical Evil and Moral Regeneration* 69.

<sup>60</sup> Kant argues that we should cultivate, and not repress, (some) of our natural feelings, such as sympathy, as aids to promoting more consistent and effective virtuous action. Allison argues that we require such extra moral forces "as a weapon against the propensity to evil"- Henry E Allison, *Kant's Theory of Taste* (Cambridge: Cambridge University Press, 2001) 232-3. A similar point is made by Nelson Potter, "Duties to Oneself,



heteronomously following our sensuous nature unconditionally, to a state of autonomously following it conditionally. We come into this world as needy beings and, as such, it is our needs that dominate. Only later do we learn to heed the call of the moral law and begin (at least ideally) to make the moral law the condition of our natural happiness.

This development argument has much going for it. It is able to give a simple and natural explanation for why human nature is beset so radically by evil. It is able to link Kant's reflections on radical evil to his wider anthropological writings. And it is also able to clearly explain why the following is not a problem. If it is at least possible that some humans have a good disposition (and Kant thinks it is),<sup>61</sup> then how can this be squared with the universality claim, i.e. that *all* humans have an evil disposition? The development argument neatly provides the answer. All humans *start* out from a state of evil, as all humans start out from a state of heteronomy and must develop toward autonomy. Thus, that some saintly humans, after much struggle, cultivation and development, might manage to successfully undertake a dispositional revolution, poses no difficulty for the claim that all humans *start* from a state of evil.

But can the development argument actually support the universality claim? Grimm, drawing upon the Kantian principle that 'ought implies can', argues that it is at least conceptually possible that, even if "the human condition seems to stack the cards in favour of evil", an exceptionally saintly person might still initially adopt a good disposition.<sup>62</sup> The fact that we all begin life, *as children*, concerned only with the satisfaction of our sensuous desires, does not *necessitate* the adoption of an evil disposition when, *as adults*, we begin using our practical reason to direct our actions, as the adoption of a disposition remains always a free and spontaneous act. As such, the development argument cannot absolutely rule out the possibility of a moral saint who, with the right sort of education, starts with, and retains, a good disposition. But it does have the important upshot of further explaining just why radical evil is so widespread (even if not universally) amongst human beings.

---

Motivational Internalism, and Self-Deception in Kant's Ethics," in *Kant's Metaphysics of Morals: Interpretative Essays*, ed. Mark Timmons (New York: Oxford University Press, 2002), 389.

<sup>61</sup> Kant argues that it is must be "possible to *overcome* this evil" disposition, since "it is found in the human being as acting freely", (6:37) although we may require the grace of God to accomplish such a momentous ethical task.

<sup>62</sup> Grimm, "Kant's Argument for Radical Evil," 171.

Wood offers a very different argument for radical evil. Wood explicitly rejects as implausible any a priori proof of Kant's radical evil thesis. But Wood does not hold Kant's radical evil thesis to be a simple empirical generalisation.<sup>63</sup> Rather, following Anderson-Gold, he conceives of it as an anthropological claim based on Kant's "theory of the purposive development of the human race's collective predispositions".<sup>64</sup> Wood argues that "Kant explicitly attributes the corruption of human nature to the *social* condition of human beings, and more specifically to the concern over comparative self-worth that characterizes people whenever they live in proximity to one another".<sup>65</sup> Radical evil becomes for Wood the misuse of our predisposition to humanity. Wood thus *equates* Kant's radical evil and unsocial sociability theses. For Wood "unsocial sociability, self-conceit, or the radical propensity to evil in human nature," are simply "three different names for the same reality".<sup>66</sup>

Kant's unsocial sociability thesis is the claim that humans have a natural predisposition to sociality, that is, a need to engage socially with other humans. This drive leads to a need to compare oneself with others, which is both a competitive spur to scientific, political and cultural progress, and also the source of much evil, in that it leads to the social vices of envy, ingratitude, malice, greed and a desire to gain ever more power and status.<sup>67</sup> By making our happiness conditional on our perceived social status, we do nothing but plunge ourselves into a state of social anxiety that ultimately precludes the achievement of the happiness that we seek. This engulfs us in a web of unjust social intrigue that ultimately undermines our worthiness to be happy. Kant's unsocial sociability thesis thus explains why immorality is so widespread and its pull so naturally strong. It also helps explain why we imprudently persist in life projects that undermine our ability to be happy, and why the likes of envy, malice, greed and a desire for power, can lead us to perpetrate evils that are not even in our self-interest.

But Kant's radical evil and unsocial sociability theses are not, as Wood thinks, *equivalent*, but are in fact *complementary*. This is in part a consequence of the fact that these two theses work on fundamentally different levels. This is made clear by Kant's

---

<sup>63</sup> Wood thus comes to reject his earlier view expressed in Allen Wood, *Kant's Moral Religion* (Ithaca, N.Y.: Cornell University Press, 1970) 219-26. See Wood, *Kant's Ethical Thought* 287.

<sup>64</sup> Wood, *Kant's Ethical Thought* 287.

<sup>65</sup> *Ibid.* 288.

<sup>66</sup> *Ibid.* 334.

<sup>67</sup> See *Ibid.* 212-5, 86-90.

discussion of the stoics, in which he expresses his dissatisfaction with their answer to the question of 'why people do evil', namely, because their inclinations tempt them. (6:59) Kant's dissatisfaction stems not from the fact that he thinks the Stoics get it wrong altogether, indeed he agrees with much of what they say, but rather from the fact that the answer they give is not *deep* enough. Kant's radical evil thesis addresses the deeper question of *why* we are the types of beings who are able to transgress the moral law *in the first place*. Kant thus wishes to know *why* our sociability turns unsocial. His answer is, of course, because of the propensity to evil rooted in the human species. Our unsocial sociability is the source of a strong incentive to adopt all sorts of *lower-order* evil maxims, but our radical evil is the embodiment of our choice of a supreme *dispositional* maxim. In other words, our radical evil *grounds*, but is not equivalent to, our unsocial sociability.<sup>68</sup>

Hence, our unsocial sociability would not even be *possible* if we did not already possess an evil disposition, although possessing an evil disposition does not in itself imply the necessity for our sociability to be unsocial. This is because, from the mere fact that we have a corrupt disposition, it need not follow that evil be so widespread, only that it be always possible. Kant's radical evil thesis tells us only that we all have our price. Kant's unsocial sociability thesis tells us why that price is, for most of us, so very low. But the unsocial sociability thesis cannot support the universality claim, for it is still possible that some saintly human may never choose to adopt a single maxim on the basis of their natural unsocial tendencies, even if it does strengthen the claim that radical evil is (at least) very widespread.

## 6. Conclusion

While Kant's proof for the universality of radical evil fails, Kant's arguments for the pervasiveness of radical evil are successful. As such, the logical defeat of the universality claim does not significantly impair the moral thrust of Kant's overall argument. There may be a few moral saints among us but, by and large, to be *human* is to be morally frail and impure (although arguably not perverse). For Kant the proper response to such a state is twofold: first, to be somewhat stern with ourselves in the pursuit of moral self-

---

<sup>68</sup> Allison criticises Wood's 'reductionism', but Wood's error lies not in being too reductive, but in *conflating* what are in fact two distinct theses - see Allison, "On the Very Idea of a Propensity to Evil," 345. Grimm makes the similar point that our unsocial sociability is not the basis but "one of its [i.e. radical evil's] consequences" - Grimm, "Kant's Argument for Radical Evil," 167.

perfection; and second, to be *humane* in our dealings with others, by reacting sympathetically to the corrupt state that is humanity.<sup>69</sup> Even a human moral saint is still no angel, and the possibility of 'falling back' into evil (via a reverse dispositional revolution) remains always a very human possibility. As such, even for the very best of us, moral complacency and self-righteousness is to be avoided, for frailty and impurity are dangers we all potentially face. As humans, evil is beyond none of us.<sup>70</sup>

---

<sup>69</sup> See the Collins notes in Kant, *Lectures on Ethics* 27:295-6.

<sup>70</sup> Thanks to Marguerite La Caze and to the journal's anonymous reviewer for helpful comments on an earlier version of this paper.