

Kant's Argument for Radical Evil

Stephen R. Grimm

Kant's doctrine of radical evil – which holds that human beings, as a species, possess an innate propensity to evil – has long been viewed as a scandal to his admirers and a stumbling block to scholars trying to piece together his argument in favor of the claim. To his admirers, the scandal stems from Kant's apparent endorsement of the Christian view of original sin, with all of its allegedly misanthropic consequences. To scholars, the stumbling block comes from the indecisive way in which Kant attempts to establish the doctrine. For example, just at the point in the *Religion within the Limits of Reason Alone* where we would expect Kant to provide a deduction supporting his view, he changes course and claims, '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' (R 6: 33).¹ Indeed, the alleged formal proof is not just 'spared' in the sense of postponed; it is 'spared' in the sense of left out of the text entirely.

To many, this indecisiveness has suggested a general lack of confidence in the view on Kant's part, and has led to a willingness to regard his treatment of radical evil as an odd, perhaps even neglectable, exception to his overall ethical project. Paul Guyer is typical: 'In *Religion within the Limits of Reason Alone* [Kant] seems to go too far by assuming that evil-doing is not just *possible* but even necessary. . . . This doctrine hardly follows from Kant's previous argument, and seems instead to rest on an odd mixture of empirical evidence and the lingering grip of the Christian doctrine of original sin.'²

In this paper, I will argue that despite the indecisiveness of Kant's argument, the doctrine of radical evil is in fact consistent with – and indeed is perhaps a natural extension of – his established views on human freedom, the moral law, and moral culpability. To make this case, I will attempt to show that the doctrine of radical evil is grounded in what (following his own usage) I will refer to as Kant's *anthropological* analysis of the human person (see, e.g., R 6: 26). By an anthropological analysis, I mean Kant's account of the different capacities human beings possess – e.g., their capacity for reason, their capacity to be affected by bodily needs and inclinations, and so on – as well as the ways in which these capacities develop in human beings over the course of their lives (MM 6: 217). It is only in light of these anthropological facts, I will argue, that the doctrine of radical evil can be properly understood.

Despite certain well-known passages in the *Groundwork* in which Kant groups together 'anthropological' and 'empirical' approaches,³ before proceeding it is important to recognize that in the majority of his work Kant carefully distinguishes

the two. As he claims in his *Anthropology from a Pragmatic Point of View* (a set of lectures which he delivered regularly for nearly thirty years), the two approaches differ in that while an anthropological analysis *begins* with empirical observations of human beings and their behavior, it goes beyond such observations by organizing them in a systematic manner – in the process raising them to the level of a genuine science (A 7: 121).⁴ Thus an anthropologist in Kant's sense of the term does not just count or observe the number of times a human being performs some action A, but instead takes the results of such observations and attempts to consolidate them in terms of a more general explanation. Put differently, the anthropologist in a sense takes for granted the fact *that* human beings perform A, and instead tries to show how the fact that they perform A points to important and abiding truths about their natures.

In order to make the case in favor of an anthropological reading, I will first try to clarify what is at stake in Kant's discussion of radical evil. In particular, I will consider the sorts of paradoxes or puzzles which are generated by Kant's claim that human beings possess an innate propensity to evil. Next, I will argue that based on the nature of the puzzles generated by Kant's account, we can rule out the claim that his view is based *merely* on empirical observation. Finally, and at greatest length, I will consider Allen Wood's important recent treatment of the role of radical evil in Kant's thought.⁵ Although I strongly agree with Wood that Kant's case for radical evil is anthropological at heart, I will argue that Wood misrepresents the true anthropological source of evil in Kant's thought. Wood takes Kant's argument to be essentially a restatement of the Rousseauian doctrine of *amour propre*, a doctrine which holds that human beings are by nature good and only come to be corrupted by their social interaction with others.⁶ Against Wood, I will try to show that this Rousseauian explanation cannot be considered Kant's because the origin of radical evil in each human being 'predates,' as it were, our interaction with others. The source of our propensity to evil does not derive from our social relationships but rather from our *composite nature* as human beings, that is, our nature as beings possessing *both* animal inclinations as well as the capacity to grasp the moral law through reason.

1. Innate Guilt

Kant begins the *Religion* by observing that, "The world lieth in evil" is a complaint as old as history' (R 6: 18), and goes on to note, with obvious regret, that the common testimony of the world provides evidence of a natural inclination in human beings towards moral corruption.

But how can we make sense of this claim? Though the empirical facts suggest some sort of natural inclination to evil, there is reason to wonder whether a notion such as 'natural evil' is even coherent. According to our common-sense notion of justice, for instance, which Kant endorses, in order for a particular fault or evil act to be imputable to a person it is necessary for the person to have freely chosen that act. In particular, to use Kant's terminology, it is necessary that at

some point the person have freely adopted a fundamental maxim, or rule-like principle, whereby she decided to act for the sake of her happiness rather than for the sake of the moral law (R 6: 36).⁷

But if this is right, then the very notion that someone could have been born with an evil maxim, in the sense of having biologically inherited it, appears unintelligible (MM 6: 382). To the extent that someone's will is caused or determined by something external to her, Kant insists, her will is no longer free; and if not free, then not a subject of moral evaluation and *a fortiori* not evil. Kant thus summarily rejects the traditional account of original sin associated with Augustine:

Whatever the nature, however, of the origin of moral evil in the human being, of all the ways of representing its spread and propagation through the members of our species and in all generations, the most inappropriate is surely to imagine it as having come to us by way of *inheritance* from our first parents; for then we could say of moral evil exactly what the poet says of the good: *genus et proavos, et quod non fecimus ipsi, vix ex nostra puto*.⁸ (R 6: 40)

According to Kant, a new theory is therefore needed to reconcile the overwhelming empirical record of human evil with the idea that imputable actions must be free and unconditioned. In order to solve this puzzle, which we might call *the puzzle of innate guilt*, Kant appeals to his prior work in metaphysics and ethics.

As we learn from the first *Critique*, the natural world – the world in time – is entirely determined by the laws of cause and effect. In the second *Critique*, however, we also learn that this cannot be the whole story. In addition to our natural or sensuous needs and inclinations, Kant insists that there exists within each of us a pure demand to conform our wills to the moral law and to do our duty even in the face of our sensuous desires. In order for the moral law to be possible, Kant therefore argues that we must possess a capacity for free choice which is undetermined by the causal events in the world. Since everything in the flow of time is determined by causal laws, however, it follows from this that our capacity for freedom must in some way be *outside* of time (C1 A551/B579). In other words, in order for the moral law to be possible, the events within time, and in particular our sensuous desires, must have no determining influence on our free choices (C2 5: 97).

If our free choices are unconditioned by natural causes and hence outside of time in the manner described by Kant, though, then the solution to the puzzle of innate guilt would appear to be within his grasp. For if we accept the idea that the will's freedom is in an important sense undetermined, then the incoherencies of a caused or determined free choice seem to disappear. Moreover, because the will must be postulated as outside of time, Kant thinks it is possible to say of the will's original propensity towards good or evil that it is, in fact, *innate* – but innate now in the special sense that it is not conditioned by events in time. As Kant puts it:

Since the first ground of the adoption of maxims, which must itself again lie in the free power of choice, cannot be any fact possibly given in experience, the good or the evil in the human being is said to be innate . . . only *in the sense* that it is posited as the ground antecedent to every use of freedom given in experience (from the earliest use as far back as birth) and is thus represented as present in the human being at the moment of birth – not that the birth itself is its cause. (R 6: 22)

This passage calls for a few clarifications. To begin with, Kant's claim that our original maxim (whether for good or evil) can be thought of as present in us from birth, as well as his talk of the maxim being 'antecedent to' our free actions in time, misleadingly suggests that there was a precise temporal moment when the maxim was chosen.⁹ But if we are to take seriously Kant's claim that the will is outside of time, then temporal categories such as 'antecedent' would be inappropriate here. As Kant later claims more clearly, what is crucial to the idea is not that the act of choosing the maxim is prior to particular human actions *in time*, but rather that it is prior to these actions *in reason* (R 6: 39). Of course, to say that these noumenal acts of freedom have a sort of rational (or logical) priority over phenomenal actions, as opposed to a temporal priority, hardly helps to make the idea of a freedom outside of time more palatable or easier to comprehend. But, as we shall later see, if nothing else this move helps to discredit approaches which regard the propensity to evil as present from birth, approaches that for Kant lead to a distorted view of the significance of radical evil.

Kant's analysis of freedom therefore appears to offer a workable solution to the puzzle of innate guilt. The maxims we adopt are imputable to us because they are entirely the product of our spontaneous free choice, and thus are not subject to any foreign influence. At the same time, though, these maxims can be thought of as innate or natural because (in the qualified way discussed above) they act as the antecedent ground for any of our particular free choices made in the course of time. Before we make our first significant free act in the natural world, according to Kant, our wills in a sense thus 'already' possess a natural propensity of some sort, either for good or evil. Unfortunately for the human race, Kant thinks that the propensity we all share is for the latter of these two alternatives: one and all, our wills opt for the evil maxim.

2. Universal Evil

This brings us to a second and arguably more troublesome puzzle raised by Kant's account of radical evil:¹⁰ If, as the results of the first puzzle seem to have shown, our wills are in fact entirely unconditioned, how can we make sense of the claim that every human being, without exception, opts for the evil maxim? In order to preserve a robust sense of freedom, a freedom which is capable of adhering to the moral law, we noticed that Kant needs to postulate a capacity for choice capable of 'absolute spontaneity' (R 6: 24), in which every act of freedom is entirely undetermined.¹¹ But

if we interpret the notion of the absolute spontaneity of the will according to its plain sense, then it seems almost unfathomable to suppose that every single human being could *just happen* to choose the same option. It would be as if someone flipped a coin 4 billion times, and every time the coin *just happened* to land heads. Eventually, any sane person would begin to suspect that such an outcome was not so entirely unconditioned after all.

Yet this is precisely what Kant wants to argue. In particular, he wants to argue that, although we possess an original *predisposition* (*Anlage*) to goodness, at the decisive juncture in our lives when we first exercise our freedom and adopt our fundamental maxim,¹² human beings as a group choose the evil maxim and thus 'acquire' a *propensity* (*Hang*) to evil.¹³ But, again, how can we make sense of the fact that human beings apparently decide, time after time, in favor of the evil maxim, and thereby acquire a propensity to evil in all their subsequent actions? Call this second difficulty *the puzzle of universal evil*.

A. An Empirical Basis?

At first glance, in the *Religion* Kant appears to offer what we might call an 'empirical' solution to the puzzle of universal evil. This interpretation of the argument is attractive because throughout the text Kant insists that in order to become convinced of humanity's innate propensity to evil, all we need to do is open our eyes and look around: from the weak to the mighty, all human beings seem to be in the thrall of sin (e.g., R 6: 20, R 6: 33).¹⁴ There is thus reason to think that according to Kant it is not necessary to show, through some form of subtle argument, that it is *possible* for human beings to possess an innate propensity to evil; rather, once we see that this is *actually* the case – once experience has shown us that *in fact* every human being, without exception, possesses a propensity to evil – then there is no further need to show how this is possible, for of course if P is actual then we can be sure that P is possible too. Thus commentators such as Philip Quinn are led to conclude:

Humans are, therefore, by nature morally evil only in the sense that the morally evil propensity to evil is, as far as we can tell on the basis of the empirical evidence available to us, universal among mankind. As Kant puts it, given what we know of humans through experience, we may presuppose moral evil to be subjectively necessary to every human, even the best of them.¹⁵

But while our experience of human evil no doubt represents what we might think of as our first *indication* of the corruption of human nature, there are a number of reasons to hold that our empirical observation stands in need of some sort of extra-empirical, background theory if the doctrine of radical evil is to have the wide scope Kant claims.

(1) In the well-known passage quoted at the outset, in which Kant claims that

we can 'spare' ourselves a formal proof of radical evil because the evidence in favor of the doctrine is so empirically obvious, Kant plainly implies that some sort of formal proof of universal evil *is* in fact available – even if, for reasons of convenience, we can ignore it on occasion.

(2) More importantly, we learn elsewhere from the *Religion* itself that we have essentially no empirical access to the maxims which guide a person's actions: 'Now through experience we can indeed notice unlawful actions. . . . But we cannot observe maxims, we cannot do so unproblematically even within ourselves; hence, the judgment that an agent is an evil human being cannot reliably be based on experience' (R 6: 20; cf. G 4: 407). But if we have no empirical access to maxims, then it is difficult to see how observation alone could serve as the basis of the doctrine of radical evil.

(3) Finally, even if it were the case that we could accurately gauge a person's internal maxims, and even if it happened to be the case that every maxim we observed was corrupt, still, all that would follow from this analysis would be that human corruption is *widespread*, not that it is actually *universal*. Since Kant is clearly interested in establishing the latter, stronger claim, however, we can conclude that Kant must think the doctrine of universal evil has more than merely empirical support.

B. An Anthropological Approach?

Recently, commentators such as Allen Wood and Robert Louden have drawn attention to the prominent role of anthropology in Kant's ethical thought.¹⁶ Whatever the status of this claim with respect to other aspects of his thought, it seems clear that a knowledge of Kant's anthropological views is essential to understanding his solution to the puzzle of universal evil. In the following crucial passage in the *Religion* he explains the need for such an approach:

That by the 'human being' of whom we say that he is good or evil by nature we are entitled to understand not individuals . . . 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 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. (R 6: 26; underline my emphasis)

In section 3, I will attempt to spell out the sort of anthropological basis for universal evil that I think Kant has in mind here. In the remainder of this section, however, I will focus on what I take to be the shortcomings of Allen Wood's alternative anthropological approach.

According to Wood, the doctrine of radical evil needs to be understood within the context of Kant's views on humanity's teleological progress over the course of

history. Further, since he thinks Kant treats history as 'a branch of biology' (208), it follows for Wood that Kant's anthropological approach is really an attempt to 'naturalize' (i.e., express in biological terms) the doctrine of universal evil. As Wood writes, 'The doctrine of radical evil is anthropological, not theological. Its basis is not religious authority but naturalistic anthropology' (291).

The basic idea behind Wood's thesis is that nature (or, more precisely, the teleological drive within nature) attempts to lift humanity beyond its original, animal state and towards a more enlightened, distinctively human state by means of the mechanism of radical evil. Thus Wood claims, for instance, that although on Kant's view nature prompts us to abandon our animal state and enter into society with one another, nature has also 'designed' human beings in such a way that, as soon as they interact with others of their species, their desire to dominate one another and to prefer their own happiness to that of their fellows immediately emerges. Wood therefore takes Kant to hold, along with Rousseau, that through social contact with other human beings, and only through such contact, evil emerges in the world.¹⁷

As Wood reads Kant, then, human beings are originally moderate or content; only later, when they begin to compete socially with others, does the propensity to evil emerge. Wood summarizes his position as follows:

So, first, the *Religion* confirms the anthropological *content* of evil, which focuses on the vices of human competitiveness. But second, the anthropological reading of the doctrine of radical evil also implies that evil has its *source* in social comparisons and antagonisms. Does the text of the *Religion* confirm the anthropological reading on this second point, too? Yes, it does. 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. Considered in abstraction from the effects of society, the natural desires of human beings are moderate, and they are disposed to contentment. What disturbs the contentment is the human being's fear of his own worthlessness in the sight of others. (288)

The Kantian source of universal evil, as Wood sees things, can therefore be traced to what Kant calls our 'unsociable sociability' (*ungesellige Geselligkeit*) (R 6: 27). We naturally desire to interact with other human beings, but this interaction in turn gives rise to feeling of envy and competitive ill-will, ultimately distorting our naturally good predispositions.¹⁸

In order to evaluate Wood's claim, we should begin by noting that Kant clearly *does* think humans are inclined to unsociable sociability, and further that he thinks this trait – especially as manifested in our incessant need to compare our situation with the situations of others – is one of the most distinctive aspects of our evil behavior. Yet, from the fact that Kant finds the phenomenon of unsocial sociability important, it of course does not automatically follow that this trait should be thought of as the *source* of radical evil in human beings. Clearly there

remains the possibility that some *other* factor might be the source of evil for Kant, with the trait of unsociable sociability in turn arising from it. And, in fact, I would like to argue that this is how Kant should be read: *contra* Wood, the phenomenon of unsociable sociability should not be thought of as the basis of the evil shared by all human beings, but rather as one of its consequences.

Perhaps the best way to resolve this issue is to turn to some of Kant's discussions of unsociable sociability. The strongest text in Wood's favor is the following:

Envy, tyranny, greed, and the malignant inclinations associated with these, assail his [i.e., mankind's] nature, which on its own is undemanding, *as soon as he is among human beings*; nor is it even necessary to presuppose that these are sunk into evil and are examples to lead him astray; it suffices that they are there, that they surround him, and that they are human beings, and they will mutually corrupt one another's moral predisposition and make one another evil. (R 6: 93; quoted in Wood: 289)

This discussion is eye-catching because it seems to assert, in plain prose, that social contact with other human beings *makes* a previously 'undemanding' human being evil. If this does not express Kant's views on the origin of radical evil, we might ask along with Wood, what possibly could?

Despite the apparent finality of this passage, however, such a reading conflicts with tenets which I suggest are yet more fundamental to Kant's thought. Consider, for instance, a similar passage from his *Lectures on Ethics* in which Kant illustrates the notion of unsociable sociability by means of the following (intentionally lighthearted?) example:

The greatest source of happiness or unhappiness, of faring well or ill, of content or discontent, lies in the relationship to other people. For if everyone alike in the town is eating rotten cheese, I eat it too, with satisfaction and a cheerful mind, whereas if everyone else were well-fed, and I alone in sorry circumstances, I would deem it a misfortune. Thus all happiness or unhappiness depends on ourselves, and on the way our minds accept the situation. (LE 27: 367)

This discussion agrees with the previous one in asserting that in isolation human beings appear undemanding and prone to contentment; in isolation, as it were, we take the rotten cheese life hands us and eat it with pleasure. Moreover, in keeping with Wood's reading this passage seems to suggest that it is only later, when we compare our situation in life with the situations of our fellow human beings, that the cheese suddenly seems to stink. And it is at this point, apparently, that evil enters into our moral character for the first time.

As a first step against such an interpretation, however, we should note that if this explanation were correct, then we would have to suppose that the man who was 'content' with his rotten cheese was not *merely* content, but was in fact *morally*

good. In particular, in keeping with the principles of Kant's ethics, we would have to suppose that this isolated person, outside of society, always acted according to a fundamental maxim formed solely out of respect for the moral law. But what reason do we have for thinking that such a person did in fact always act out of respect for the moral law? None, that I can see. Indeed, the more likely assumption is that he, like all of the other human beings we have observed, was naturally inclined to act *primarily out of a concern for his own happiness* – and only secondarily, if at all, out of a concern for the moral law. In short, before entering into society, presumably the maxim which grounded his actions was something like the following: pursue that which pleases me or makes me content, and avoid that which pains me. But according to Kant this is this maxim which represents the very definition of evil! *Prima facie*, therefore, Kant seems committed to the view that the man was evil *before* it even occurred to him to compare his position with that of his fellows, and hence that he was evil before his unsociable sociability emerged.¹⁹

Some background is needed to appreciate the force of this point. Most importantly, there is the idea that Kant lays out and endorses at the very outset of the *Religion*, which he names 'rigorism'.²⁰ According to rigorism, a human being is *either* morally good *or* morally evil. 'Experience' suggests to us that there is a third, middle position available here, namely, that human beings are a mixture of good and evil. But Kant insists that experience is mistaken. The moment that we fail to act according to a subjective maxim committed to the moral law, he argues, we do not become slightly less good, as if we just temporarily lapsed from the good maxim; rather, what this lapse shows is that we were *never committed to the moral law to begin with*. Put differently, it shows that the maxim which we had in fact endorsed all along was along the lines of the one we ascribed to the cheese eater above: I will try to do my duty, yes, but only so long as it is convenient for me. But, to repeat, according to Kant this maxim is not just *partially* good or *incompletely* good. It is, rather, downright evil. He writes:

If the law fails nevertheless to determine somebody's free power of choice with respect to an action relating to it, an incentive opposed to it must have influence on the power of choice of the human being in question; and since, by hypothesis, this can only happen because the human being incorporates the incentive (and consequently also the deviation from the moral law) into his maxim (in which case he is an evil human being), it follows that his disposition as regards the moral law is never indifferent (never neither good nor bad). (R 6: 24; underline my emphasis)

Hence in order to classify the isolated 'contented' man Wood speaks of as good, according to Kant we have to suppose that his fundamental maxim was such that all of his actions outside of society were performed out of a respect for the moral law, and only out of a respect for the moral law. But since such a man, like the rest of us, would of course have been influenced by sensuous incentives, it seems virtually impossible to suppose that such a man would have been good – at least good in the essentially unimpeachable sense required by Kant's rigorism.

There are thus *prima facie* reasons to deny that unsociable sociability could serve as the original source of radical evil or account for its universal scope. In what follows, I will try to show that Kant offers *another* anthropological argument for radical evil, different from the one emphasized by Wood, which better accords with Kant's overall project. Moreover, I will argue that the reading which I offer can accommodate Wood's interpretation, and therefore offers a more complete explanation.

3. Another Anthropological Approach

In the *Religion* Kant claims that human beings have three predispositions (*Anlagen*) to good: a *predisposition to animality*, which aims at self-preservation, sexual reproduction, and community with other human beings; a *predisposition to humanity*, which prompts human beings to compare their own worth with the worth of others; and finally a *predisposition to personality*, which among other things makes a person capable of appreciating the moral law. For our purposes the first two of these predispositions are the most noteworthy because, according to Kant, these are the two which can be used inappropriately. Kant explains:

By the predispositions of a being we understand the constituent parts required for it as well as the forms of their combination that make for such a being. They are *original* if they belong with necessity to the possibility of this being, but *contingent* if the being in question is possible in itself also without them. (R 6: 28)

These remarks are significant because they indicate that the predispositions Kant identifies are precisely the sort of things which will enable us to amplify (to use the language of the first *Critique*) our understanding of what is contained *a priori* in the concept of 'human being'. This, I take it, is what Kant means when he claims that such predispositions are 'original' to human beings, but at the same time 'contingent' in that they are not contained explicitly (and hence analytically) in the very definition of what it means to be a human being. Thus, if these predispositions are such that they naturally contribute to our propensity to evil, then we are entitled to say that we have found a characteristic which is common to all human beings and hence 'universal', while at the same time not directly *necessitating* this evil and hence robbing us of our freedom. And this is precisely what we have been searching for: a way to account for the universal scope of radical evil which is nevertheless imputable to us.

The first thing which bears emphasizing about these predispositions – especially the predisposition to animality – is that Kant thinks that 'considered in themselves' they are 'good' (R 6: 58). Unlike the Stoics, Kant therefore claims that it is not the mere fact that we have animal needs and inclinations which gives rise to evil, but rather the improper priority we grant to these needs and inclinations. But again, we want to know, why does the human will choose to make our

natural inclinations primary? Why does it choose to make our fundamental maxims tied to their satisfaction? And why are such choices universal? In his *Lectures on Religion*, in which he discusses the result of an initially merely 'animal' human being coming to discover his or her capacity for reason, Kant sheds important light on these questions:

This predisposition to good, which God has placed in the human being, must be developed by the human being himself before the good can make its appearance. But since at the same time the human being has many instincts belonging to animality, and since he has to have them if he is to continue being human, the strength of his instincts will beguile him and he will abandon himself to them, and *thus arises evil*, or rather, when the human being begins to use his reason, he falls into foolishness. *A special germ towards evil cannot be thought, but rather the first development of our reason toward the good is the origin of evil.* And that remainder of uncultivatedness in the progress of culture is again evil. (LR 28: 1078)

He continues shortly thereafter:

Hence [a human being] finds evil *first* when his reason has developed itself far enough that he recognizes his obligations. St. Paul says that sin follows upon the law. . . . As soon as the human being recognizes his obligation to the good and yet does evil, then he is *worthy of punishment*, because he could have overcome his instincts. And even the instincts are placed in him *for the good*; but that he exaggerates them is his own fault, not God's. . . . This *justifies* God's holiness, because by following this path the whole species of the human race will finally attain to perfection. But if we ask where the evil in individual human beings comes from, the answer is that it exists on account of the limits necessary to every creature. (LR 28: 1079; underline my emphasis)

I take the picture of human development in these passages to be as follows. In the early stages of life, the animal element of our nature basically determines our behavior. Our desire to preserve our lives, as well as our desire to pursue pleasure or happiness and avoid pain or unhappiness, compel our actions. Further, in the early stages of development these desires are not just amoral or partly moral; rather, they are *good* – in the sense that they seem to be doing nothing more than playing the proper role which nature has assigned to them. As human beings continue to mature, however, momentous new capacities begin to emerge. Most significantly, of course, our capacity for *reason* begins to emerge.²¹

With the onset of reason, the human moral landscape is fundamentally altered. Whereas prior to reason we were guiltless in our pursuit of pleasure or happiness, with reason's emergence our sudden awareness of the moral law forces us to view these desires in an entirely different light. In particular, when reason asserts itself we discover, to our dismay, that the desire to seek our own pleasure

or happiness is in fact a corrupt goal in the eyes of the moral law; against such desires, reason teaches us that we are only truly free, and indeed truly good, when we act out of respect for the law alone.

So where does evil enter into this picture? From the previous passage we learn that Kant endorses the spirit of the Pauline claim that *sin follows upon the law*. Put into a Kantian idiom, the idea seems to be that when we learn what we *ought* to be doing (obeying the moral law), our natural needs and inclinations (which a person 'has to have if he is to continue being human') suddenly take on a rebellious and, for the first time, an evil role. The law comes to us all as an instrument of freedom and enlightenment, but, as our initial act of freedom, we choose to reject the law in favor of the needs and inclinations which were a natural part of our identity before the coming of the law.

It is in this way, I would like to suggest, that Kant attempts to solve the puzzle of universal evil. In short, his claim is that by the time reason emerges in the human agent – and with reason the moral law – our natural needs and inclinations have already acquired a kind of foothold within the soul which reason finds is nearly impossible to dislodge. Hence, by noting that we are composite beings, part animal and part rational, and moreover by noting that the rational element emerges later in our development and is forced to confront a suddenly rebellious set of animal needs and inclinations, Kant thinks he can explain why all of us *freely* adopt evil maxims, i.e., maxims according to which the demands of the moral law must first accommodate themselves to the demands of our established and clamorous animal inclinations.

It should be clear at this point that Kant's resolution of the puzzle of universal evil is not entirely satisfying. Granted, through his 'anthropological research' into the composite nature of the human being, as well as his study of the complex process of human development, Kant has shown how the human condition is such that the possibility for evil awaits us as soon as we reach the age of reason. But what he has not shown is why all human beings, one after another, actually choose to give priority to the established sensual inclinations in the soul. We want to ask: Even if the human condition seems to stack the cards in favor of evil, doesn't the bedrock *ought-implies-can* principle imply that the odd person might always come along and freely chose to obey the moral law from a sense of duty as the law requires? And wouldn't the mere existence of this possibility be enough to undermine the doctrine of universal evil?

At this point, I think it is fair to say that Kant acknowledges, with conspicuous humility, that he has run out of answers.²² The ultimate cause of evil for Kant, and the reason why we freely choose to subordinate the moral law to our animal inclinations time after time, lies hidden in the dark depths of the human heart. On the final analysis, he thus concedes that the cause of evil is 'inscrutable' (R 6: 21), 'incomprehensible' (R 6: 44), 'unfathomable' (R 6: 60), and 'forever shrouded in darkness' (R 6: 59). In describing the anthropological conditions which make evil such an attractive option for human beings, Kant thus offers a *partial* solution to the puzzle of universal evil. The *complete* solution to the puzzle, he insists, lies beyond the reach of our intellect.²³

4. Wood Again

Before concluding, we first need to return briefly to Allen Wood's important alternative account of mankind's propensity to evil. We saw that, as Wood reads Kant, the propensity to evil first arises when a person enters into society and begins to compare his own status with that of his neighbors. Moreover, we noted that this idea of unsociable sociability is indeed prominently featured in the text: especially the passage from the *Religion* (quoted above in section 2.b) in which Kant claims that human beings in isolation are content and unbothered, and that it is therefore only the society of other people which 'makes' human beings evil (*sich einander böse zu machen*).

Against Wood's account, I have argued that the source of evil is instead connected with our composite nature as human beings, and that therefore even the human being 'in isolation,' before his entrance into society, would still be considered evil on Kant's view. But then, if this is true, what Kant describes as our 'first' turn to evil (i.e., our adoption of the evil maxim) cannot be attributed to our entrance into society; rather, it has its source in the conflict which arises within every human being when (at a certain stage in our life) the incipient faculty of reason confronts our entrenched animal needs and inclinations.

So which explanation best captures Kant's view? One option, of course, would just be to conclude that Kant carelessly provided two mutually inconsistent explanations; his admiration for Rousseau, perhaps, led him to try to accommodate the idea of unsociable sociability even though it did not quite harmonize with his considered account. Overall, however, the texts support a resolution of the following sort: Although it is the conflict between reason and our animal desires which *first* gives rise to evil (and thus can properly be thought of as the *source* of our propensity to evil), our social interaction with other people brings this evil to an entirely different level of sophistication and corruption. Consider, for instance, Kant's description of the kinds of vices which the 'predisposition to humanity' (i.e., our predisposition to enter into society with others) introduces:

These vices [of jealousy and rivalry], however, do not really issue from nature as their root but are rather inclinations, in the face of the anxious endeavor of others to attain a hateful superiority over us, to procure it for ourselves over them for the sake of security, as preventive measure; for nature itself wanted to use the idea of such a competitiveness (which in itself does not exclude reciprocal love) as only an incentive to culture. Hence the vices that are grafted upon this inclination can also be called vices of culture, and in their maximum degree of malignancy (where they are simply the idea of a maximum of evil that surpasses humanity), e.g., in *envy, ingratitude, joy in others' misfortunes*, etc., they are called *diabolical vices*. (R 6: 27; underline my emphasis)

The central idea of this passage seems to be not that social relations are the root of evil or vice, but rather that they give rise to a *special sort* (or subset) of vice: the 'vices of culture'.

At the risk of being repetitious, we are now in a position to see how a passage quoted earlier (in section 3) also supports this interpretation:

This predisposition to good, which God has placed in the human being, must be developed by the human being himself before the good can make its appearance. But since at the same time the human being has many instincts belonging to animality, and since he has to have them if he is to continue being human, the strength of his instincts will beguile him and he will abandon himself to them, and *thus arises evil*, or rather, when the human being begins to use his reason, he falls into foolishness. *A special germ towards evil cannot be thought, but rather the first development of our reason toward the good is the origin of evil. And that remainder of uncultivatedness in the progress of culture is again evil.* (LR 28: 1078; underline my emphasis)

Once more, the idea here seems to be that the propensity to evil arises from the clash between reason and our animal desires, but that our immersion into society or culture adds another, perhaps even more diabolical, dimension to evil. We are evil due to the fact that when reason takes shape in us we subordinate its directives to the inclinations of our animal nature, and when we subsequently turn to confront our fellow human beings this tendency to favor our own happiness and pleasure gives rise to a whole new arena for vice. It is in this sense that society/culture 'adds to' our evil, or in Kant's words makes us '*again evil*'. But then, *contra* Wood, it should now be clear that society does not *make* us evil simpliciter, as it were – in the sense that it first gives rise to evil. It is our composite nature, and not our entry into society, which has this dubious distinction for Kant.

5. Conclusion

I have tried to show that in order to understand Kant's doctrine of radical evil we need to consider his anthropological analysis of the human person. In particular, we need to see how Kant thinks our composite human nature – as subject to both sensual inclinations and rational imperatives – creates a fertile ground for rebellion and sin which we all, for reasons ultimately mysterious to even Kant, choose to partake in. Moreover, I have tried to show that this account is preferable to Wood's because (a) it takes Kant's rigorism seriously, and (b) it can accommodate the importance of unsociable sociability in Kant's thought, without requiring that we recognize unsociable sociability as the source of radical evil.²⁴

Stephen R. Grimm
 Philosophy Department
 University of Notre Dame
 Notre Dame
 Indiana 46556-4619, USA
 sgrimm@nd.edu

NOTES

¹ All references to Kant's works will be included in the text, according to the following abbreviations: A=*Anthropology from a Pragmatic Point of View*, trans. and intro. Mary Gregor (The Hague: Martinus Nijhof, 1974); C1=*Critique of Pure Reason*, ed. and trans. Paul Guyer and Allen Wood (New York: Cambridge University Press, 1998); C2=*Critique of Practical Reason*, in *Practical Philosophy*, ed. and trans. Mary Gregor (New York: Cambridge University Press, 1996); G=*Groundwork of the Metaphysics of Morals*, in *Practical Philosophy*, ed. and trans. Mary Gregor (New York: Cambridge University Press, 1996); LE=*Lectures on Ethics*, ed. Peter Heath and J.B. Schneewind, trans. Peter Heath (New York: Cambridge University Press, 1997); LR=*Lectures on Religion*, in *Religion and Rational Theology*, ed. and trans. Allen Wood and George di Giovanni (New York: Cambridge University Press, 1996); MM=*The Metaphysics of Morals*, in *Practical Philosophy*, ed. and trans. Mary Gregor (New York: Cambridge University Press, 1996); R=*Religion within the Limits of Reason Alone*, in *Religion and Rational Theology*, ed. and trans. Allen Wood and George di Giovanni (New York: Cambridge University Press, 1996).

² Guyer 1998: 192. For earlier admirers, including Schiller and Goethe, who took a dim view of Kant's treatment of radical evil, see Fackenheim 1954: 339–340.

³ Only to dismiss them both as irrelevant to the moral law, e.g., G 4: 389.

⁴ For a nice discussion of Kant's treatment of anthropology as a science, as well as the distinction between empirical and anthropological approaches, see Mary Gregor's introduction to *Anthropology from a Pragmatic Point of View*.

⁵ Wood 1999. Hereafter page references to this work will be included in parentheses in the text.

⁶ Wood writes: 'Kant and Rousseau express one and the same doctrine in diametrically opposed words' (292; underline my emphasis).

⁷ Of course, Kant does not think the adoption of such a maxim needs to have been explicit. The idea is rather that when one acts 'freely' for the first time and 'with consciousness', a person's maxim is manifested in his incentives and goals (R 6: 42).

⁸ From Ovid: 'Race and ancestors, and those things which we did not make ourselves, I scarcely consider as our own'.

⁹ Christine Korsgaard, for example, seems to take Kant's claim that the evil maxim is present from birth at face value. See her 1996: 181. A more accurate analysis is given by Quinn 1990: 'Fixing on any particular time in that interval [between birth and the agent's first evil action] would seem to be wholly arbitrary, and besides the length of the interval is apt to vary from one agent to another. So postulating that the evil disposition is present from birth, if present at all, and thus is innate yields a gain in simplicity without any offsetting loss in explanatory power' (p. 420).

¹⁰ Troublesome, at least, with respect to the clarity of Kant's arguments.

¹¹ Or, at least, every act of 'original' freedom, whereby we establish our initial maxims, must be so.

¹² It is almost inevitable to speak in such quasi-temporal terms as 'decisive juncture.' But it should once more be borne in mind that the relation here is rational or logical, not temporal.

¹³ For Kant, a propensity thus differs from a predisposition in that, while both can be thought of as innate, the former is acquired whereas the latter seems to belong to the very fabric of human nature.

¹⁴ See also, in the *Anthropology*: 'The question can be raised, whether our species should be considered a *good* race or an *evil* one . . . and then I must admit that there is not

much to boast about it. If we look at man's behavior not only in ancient history but also in contemporary events, we are often tempted to take the part of *Timon* the misanthropist in our judgments' (A 7: 332).

¹⁵ Quinn 1988: 111. Interestingly, earlier in his career Allen Wood also endorsed this interpretation; see his 1970: 219–26.

¹⁶ See Wood 1999 and Loudon 2000.

¹⁷ But, Wood says, according to Kant our propensity to radical evil should not be thought of as a *flaw* in nature's 'design'; on the contrary, it is part of her subtle plan to improve and advance the human species. By first preferring our own happiness and interests to those of our fellows, nature gives us a tremendous incentive to develop our talents and capabilities as we seek to dominate others; especially, we develop our capacity to *reason*, if only at first in an instrumental way, in order to achieve our selfish ends. Once reason has appeared on the scene, however, it acquires an authority which transcends our selfish ends. Indeed, eventually we learn, by means of the moral law which reason reveals, that our fellow human beings are not simply adversaries to be dominated, but are rather ends in themselves. As human development continues still further, according to Wood, it is possible that the original propensity to radical evil, which lead to our 'discovery' of reason as we sought a competitive advantage, will melt away altogether as we approach a true kingdom of ends. Hence for Wood, 'Kant's ethical thought is fundamentally about the human race's collective, historical struggle to develop its rational faculties and then through them to combat the radical propensity to evil that alone made their development possible' (296).

¹⁸ This basically Rousseauian interpretation of radical evil is also favored by Sharon Anderson-Gold: 'The Kantian conception of moral evil as outlined in the *Religion* defines moral evil as a 'propensity,' an acquired characteristic of the faculty of freedom so fundamental that it is coeval with the social and cultural condition of mankind' (2001: 4). (See especially ch. 3 for her development of this theme.) Although I do not address Anderson-Gold's work directly here, many of the arguments I present against Wood's view would apply to her interpretation as well.

¹⁹ Admittedly, this reading presents us with a strange and perhaps unintelligible creature, the 'presocial man,' who sounds more like an *ad hoc* invention than a genuine possibility. But note that on Wood's reading unless the existence of such a man can be taken seriously – unless, that is, such a man is somehow more than an abstraction – then Wood appears to be left with the following difficulty. On the one hand, if he emphasizes that all men are inevitably raised in *some* society (which of course no one would deny), then since exposure to society on his view somehow triggers our propensity to radical evil, it would seem to follow that all human beings acquire their propensity to radical evil along with their introduction into society; that is, from birth. But it is not easy to see how such a view significantly differs from the sort of Augustinian account Kant emphatically rejects. On the other hand, however, if Wood argues that our tendency to compete with one another only arises gradually over time – that it should be thought of as some moment which occurs in the process of *socialization*, perhaps, as opposed to some moment which occurs when one is first admitted into society – then it would appear that the cheese eater I consider would indeed be a genuine possibility. The cheese eater might, for instance, be supposed to have all the social training needed in order to activate *amour propre*, yet have removed himself from society for one reason or another before it became full-fledged. I have doubts about whether this is an entirely adequate response, however, and I think more could be said to address this point. I thank Allen Wood for pressing the objection.

²⁰ For more on the central role of rigorism in Kant's ethical thought, see Allison 1990 esp. ch. 8.

²¹ For a more detailed discussion of Kant's theory of adolescent development, see Savage 1991 esp. pp. 66–67.

²² Alternatively, one might say that since on Kant's view the moral law alone furnishes a fully sufficient reason for action, it follows that he *has* to say that the origin of evil is incomprehensible because there is *no reason at all* to adopt the sort of heteronomous maxim involved in radical evil. Thus perhaps it is more accurate to conclude that Kant is coming to terms with the requirements of his own theory here, rather than simply running out of answers. I thank Kevin Robbins for this point.

²³ Although this seems to be Kant's final word, perhaps a bit more could be said on Kant's behalf in defense of the doctrine. For example, although it is true that in order for a fault to be imputable to me it must have been possible for me to avoid the fault, the relevant sense of possibility here is vague. Having been born an American citizen, it is possible for me to become President of the United States in a way that it is not possible for me to become Prime Minister of Germany; nevertheless, due to certain unfavorable circumstances (lack of financial resources, etc.) the actual possibility that I will become President is exceedingly low. We might thus call the first sense of possibility 'theoretical possibility' and the second sense 'actual possibility.' Perhaps then Kant is asking us to look at radical evil in the following way: Yes, it is *theoretically possible* for us to obey the moral law and bring our sensual inclinations into line (after all, imputability demands it), but *practically* speaking we are born into circumstances where the voice of the moral law is so muffled in relation to our clamoring desires that it is extremely unlikely we will be able to respond to its demands.

At any rate, before finally concluding that Kant's explanation of the universal scope of evil is unsatisfying, it seems that much more needs to be said about the relationship between possibility and imputability here.

²⁴ I would like to thank Karl Ameriks, Jason Baldwin, Christian Miller, Kevin Robbins, Allen Wood, and especially Phil Quinn for helpful comments on earlier versions of this paper.

REFERENCES

- Allison, Henry (1990), *Kant's Theory of Freedom*. New York: Cambridge University Press.
- Anderson-Gold, Sharon (2001), *Unnecessary Evil: History and Moral Progress in the Philosophy of Immanuel Kant*. Albany: SUNY Press.
- Fackenheim, Emil (1954), 'Kant and Radical Evil', *University of Toronto Quarterly* 23.
- Guyer, Paul (1998), 'Immanuel Kant,' in *The Routledge Encyclopedia of Philosophy*, ed. Edward Craig. New York: Routledge.
- Kant, Immanuel, *Anthropology from a Pragmatic Point of View*, trans. and intro. Mary Gregor. The Hague: Martinus Nijhof, 1974.
- *Critique of Practical Reason*, in *Practical Philosophy*, ed. and trans. Mary Gregor. New York: Cambridge University Press, 1996.
- *Critique of Pure Reason*, ed. and trans. Paul Guyer and Allen Wood. New York: Cambridge University Press, 1998.
- *Groundwork of the Metaphysics of Morals*, in *Practical Philosophy*, ed. and trans. Mary Gregor. New York: Cambridge University Press, 1996.

- *Lectures on Ethics*, ed. Peter Heath and J.B. Schneewind, trans. Peter Heath. New York: Cambridge University Press, 1997.
- *Lectures on Religion*, in *Religion and Rational Theology*, ed. and trans. Allen Wood and George di Giovanni. New York: Cambridge University Press, 1996.
- *The Metaphysics of Morals*, in *Practical Philosophy*, ed. and trans. Mary Gregor. New York: Cambridge University Press, 1996.
- *Religion within the Limits of Reason Alone*, in *Religion and Rational Theology*, ed. and trans. Allen Wood and George di Giovanni. New York: Cambridge University Press, 1996.
- Korsgaard, Christine (1996), 'Morality as Freedom', in *Creating the Kingdom of Ends*. New York: Cambridge University Press.
- Louden, Robert (2000), *Kant's Impure Ethics*. New York: Oxford University Press.
- Quinn, Philip (1988), 'In Adam's Fall, We Sinned All', *Philosophical Topics* 16.
- (1990), 'Saving Faith from Kant's Remarkable Antinomy', *Faith and Philosophy* 7.
- Savage, Denis (1991), 'Kant's Rejection of Divine Revelation and His Theory of Radical Evil', in *Kant's Philosophy of Religion Reconsidered*, ed. P. Rossi and M. Wreen. Bloomington: Indiana University Press.
- Wood, Allen (1970), *Kant's Moral Religion*. Ithaca: Cornell University Press.
- (1999), *Kant's Ethical Thought*. New York: Cambridge University Press.