Back to Article

The Consistency Of A Kantian View Of Radical Evil
by Jonathan Lipps

This is a review of Kant's Exposition of Evil in *Religion Within the Boundaries of Mere Reason.*

In Book I of his *Religion within the Boundaries of Mere Reason* (hereafter the *Religion*), Kant discusses what he calls the radical evil in human nature. The discussion proceeds in four stages, i.e., sections, concerning (I) the original predisposition to good in human nature, (II) the propensity to evil in human nature, (III) the fact that the human being is by nature evil, and (IV) the origin of evil in human nature. In this paper I will give a brief sketch of the first three sections, with an eye towards helpful exposition, and then focus primarily on the last, wherein Kant explores, given that the human being is by nature evil, what the "first cause" of said evil nature might be, and whether we can even say anything about it at all. I will maintain the thesis that Kant's position in section IV is at least consistent, given the preceding discussion, and specifically when the distinction between noumena and phenomena is taken into account.

Before we even get to section I, Kant sets the stage by arguing that a human being can be good by nature or evil by nature, but not both or neither. He comes to this conclusion because of his conception of freedom, and because of how he defines an evil person. As for that, he says, "We call a human being evil, however, not because he performs actions that are evil (contrary to law), but because these are so constituted that they allow the inference of evil maxims in him" (Ak 6:20). He goes on to explain how we can notice through experience certain evil actions, but never maxims. To judge that a human being is evil, then, would require we infer *a priori* from an evil action to an evil maxim (6:20). Since Kant immediately moves to a somewhat different part of the argument, it seems that he takes our ability to judge a human being evil as given, with the result that we simply must be able to infer *a priori* evil maxims in the way described above. Then, I would reconstruct the remainder of the argument as follows: the ground of any evil maxim is an over-arching maxim which permits the formulation of evil maxims. This maxim might

also incorporate the pursuing of good actions, but it is still evil, as long as it permits the formulation of just one evil maxim. But then, because of how we have said that this maxim is evil, we cannot suppose that an agent has a fundamental good maxim as well, because these two maxims would either contradict or the other would become evil because of the one. So, since the adoption of these maxims is imputable to the human being, it must be a result of the use of freedom, but before experience (more accurately, not in time). Therefore every human being we run across in experience is responsible in this way, whether for the adoption of the good maxim or the evil maxim.

Assuming that human beings are either good or evil by nature, it might seem odd for the first two sections of Book one to be entitled "Concerning the original predisposition to good in human nature" and "Concerning the propensity to evil in human nature", respectively, since these headings could imply that human beings are, in some sense, both good and evil. However, Kant means decidedly different (and technical) things by "predisposition" and "propensity". By the predispositions of a being, he means, "the constituent parts required for it as well as the forms of their combination that make for such a being" (6:28). A propensity, on the other hand, is a type of predisposition: one to "desire an enjoyment which, when the subject has experienced it, arouses inclination to it" (6:29). So listing out certain predispositions of a being will give us a sort of definition of that type of being, whereas the propensities of a being relate instead to the faculty of desire for that being.

The predispositions which Kant thinks determine the human being are the predispositions to (1) animality, (2) humanity, and (3) personality (6:26). What exactly these are I will not discuss here; it is more interesting that Kant thinks they are predispositions to the good. He says, in fact, that they "are not only (negatively) good (they do not resist the moral law)…but they are also predispositions to the good (they demand compliance with it)" (6:28). So, the very predispositions which define what essentially a human being is also demand compliance with the moral law. Now, all this raises the question of how it can be that the concept of a human being is good,

while all human beings are actually evil by nature (which is what Kant wants to argue later). Another way to put the question is as follows: if the definition, in terms of predispositions, of a human being is of a being bound to the moral law (i.e., the very concept of a human is of a being that should do good), how can it be that no actual human being morally fulfills this concept? But phrased this second way, the question loses some of its force, since we can see that there is no contradiction in supposing that the human being is a creature bound to the moral law, having a predisposition to regard morality as an incentive to action, yet at the same time a propensity to incorporate the moral law into her maxims, as an incentive, only as subordinated to the incentives of inclination.

We move now to section II, on the propensity to evil in human nature. Kant makes an important point immediately, which is that "a propensity can indeed be innate yet may be represented as not being such: it can be thought of…(if evil) as brought by the human being upon himself" (6:29). As we have already hinted at above, Kant would need to represent the propensity to evil in this way, since, as he says, "it is only possible as the determination of a free power of choice" (6:29). Thus the being under consideration is responsible not just for the evil acts which are a result of the propensity but for the propensity itself; if he were not, then he could hardly be responsible for the evil acts either. (Of course, as we will see later, trying to discover the origin of this propensity becomes very difficult.)

Most of section II is actually spent spelling out three different grades, or degrees, of the natural propensity to evil. First, there is frailty, or weakness of the will. Second, there is impurity, which corresponds to the human habit of shoring up the will to do good with incentives other than duty. Third, there is depravity, which corresponds to a straightforward subjugating of the moral law to other, non-moral maxims. Curiously, after simply listing and explicating these three concepts, Kant states (rather grandly), "It will be noted that the propensity to evil is here established…in the human being, even the best…" (6:30). A careful reading of the text proves fruitless in a search to find any explicit argument for which the foregoing is supposed to be the conclusion. And this conclusion is an important

one for Kant—something he realizes: "…and so it also must be if it is to be proved that the propensity to evil among human beings is universal…" (6:30). The only way to rescue Kant's statement is to assume that he believes it obvious that all humans in fact characteristically act out these three degrees of the propensity to evil. We will consider whether or not this is truly obvious when we discuss the next section.

Before continuing to section III, however, there is one more important piece of argumentation to look at in section II. Towards the end of that section Kant notes a bit of a tension in his view: "Nothing is, however, morally (i.e. imputably) evil but that which is our own deed. And yet by the concept of a propensity is understood a subjective determining ground of the power of choice that precedes every deed, and hence is itself not yet a deed" (6:31). In other words, the propensity is supposed to be that which sets up how our deeds will generally go, so it would seem to be a contradiction to call it a deed itself. But if we do not call it a deed, then we have no way of imputing it (as an action) to the subject, and therefore no real way of blaming the subject for any evil results of the propensity. Kant's solution is to say that "deed" actually has two meanings, both of which result in the requisite sort of moral responsibility. In the new meaning, "deed" can refer to an action which is the use of freedom through which the supreme maxim is adopted. This maxim is adopted, of course, before any "deeds" (in the former, standard sense) have been realized, since these deeds are one and all the result of the supreme maxim. To tell such a story is consistent, but raises a host of questions about the relationships between freedom, time, consciousness, and experience. Some relevant ones I will discuss, since they crop up again in section IV. But for now, let us just say that there is nothing contradictory in the assumption that there are two kinds of deeds. (In fact, we will see that ultimately there is nothing more that Kant could consistently claim to say, even if he wanted.)

Section III, entitled "The Human Being is by Nature Evil", has two discussions which can help us in the current enterprise of understanding Kant's views on the origin of evil. The first is of the "proof" of the universality of evil, to which we referred above. A

formal proof would consist of certain empirical evidences joined together systematically by a science of anthropology (a science which Kant views in some sense as ultimate [Wood 1], though whether or not it is attainable in pure form is questionable). Again, though, Kant does not attempt to give a formal proof of any kind, "in view of the multitude of woeful examples that the experience of human deeds parades before us" (6:33). Accordingly, he proceeds to give us a number of examples of corruption. One important question, then, is whether this admittedly a posteriori trotting-out of human maleficencies suffices to convince us of the universality of evil. Kant certainly thinks it does, but ultimately this question can be debated. I, for one, would not want to oppose Kant.

The second discussion concerns the ground of the propensity towards evil (to be distinguished, of course, from the origin thereof, which is discussed in section IV). Kant rightly notes that this ground cannot be placed in the sensuous nature of the human being (6:35). This is because we cannot be held responsible for our natural inclinations, i.e., those which do not result from a free choice. The only alternative location for the ground of the propensity towards evil would then seem to be the intelligible, i.e., noumenal, nature of the human being (if it can be correctly said that we have a noumenal nature). But it is a well-known thesis of Kant's that we can know nothing about any noumenal objects, with the result that we could never know the ground of the propensity towards evil, apart from the fact that it had something to do with the concept of the human being as a free, noumenal being. In fact, I will look at section IV in precisely this light, since I think it is the only way of reading Kant consistently there. That being said, let us move on (finally) to section IV, which has the heading, "Concerning the Origin of Evil in Human Nature".

Kant defines an "origin" as the descent of an effect from its first cause. Of course, we must take "first" here to mean something different than "first in time", since we have already seen that the evil in humans must be said to be there before the first (temporally speaking) evil action they commit, and so if we are looking for a temporally first cause, we will eventually be forced to go back in

time, previous to the birth of whichever person we are considering. But this is unintuitive, because a person can hardly be responsible for anything before she is born, and, as we keep saying, the cause of evil effects must also be imputable to the subject (that is, it must be located in the operation of the freedom of the subject), which requires that the cause itself be an action or a maxim. Moreover, it is obvious that there must be an origin of human evil in time, if we agree that there was a time before humans existed, in which human evil likewise did not exist. Kant appears not to be interested in considering evil this way (probably for the reasons of responsibility given above), and so it would be unfair to count omission of the fact that there must have been a first evil act (just as there was a first human, however we define human) against him. Let us then keep in mind that, when we are speaking of the origin of evil, the "first cause" of evil does not refer to the first cause, temporally.

Kant says as much when he notes that there are two types of origin: origin according to reason and origin according to time (6:39). In the first type, we consider only the effect's being, and in the second we consider its occurrence as well (its being-when, we might say). Here we have the first puzzle of the section, in that it is not clear how exactly we consider a thing's being only, and not its being-when. At least, we cannot imagine such an object, and Kant would argue similarly—since time is a form of intuition, anything we do intuit will necessarily be represented as in time. It seems, then, that the only way to resolve the puzzle would be to admit that if we consider only an effect's being, i.e., simply that it exists, we are considering the effect as a noumenon. This conclusion supports my strategy of phrasing Kant's discussion here as relying crucially on this distinction of Kant's between noumena and phenomena.

But if I am right, we are immediately plagued by a host of difficult questions. For instance, how do we know that the concept of an effect is even possible in the noumenal world? Is such a concept not perhaps crucially related to the concept of time? Even more troubling is the question of how we could know that a noumenal effect can be correlated with a phenomenal effect. That is, how do we know, when we are considering an effect, that it is the same object considered

merely ontologically as it is considered in time? The doctrine of restricted epistemic access to the noumenal world would seem to leave this skepticism without hope of an answer. However, that there can be no intelligible answer is precisely what we should expect from any question asked about noumenal objects, since we can know nothing about the noumenal world.

So it is surprising that, as to this second question, Kant gives the seeds of a response. He says, "[The] effect's…occurrence…as an event…is referred to its cause in time. If an effect is referred to a cause which is however bound to it according to the laws of freedom, as is the case with moral evil, then the determination of the power of choice to the production of this effect is thought as bound to its determining ground not in time but merely in the representation" (6:39). This passage implies that, just as we have two ways of looking at the event (ontologically and temporally), we have two ways of looking at its causes—one is found in the representation of reason, and one is found in a preceding state. Thus Kant could be seen as advocating some kind of (in principle unknowable) connection between noumenal effects and phenomenal events. The picture I am imagining has the free will situated in the noumenal self of the human being. The subject then performs an evil action. Considering the effect as an action, the ground of this effect is, of course, simply in the phenomenal constitution of the subject (there is no tracing it to any noumenal cause). However, considering the effect as a moral event, the ground of the effect is in the free will of the subject, i.e., the noumenal self. While it may not be clear how exactly to conceive of this distinction, I believe it gets the job done, in a logical sense—it locates the origin of moral evil somewhere outside the physical nature of the human being, which we decided above was necessary. Of course, for Kant to remain completely consistent with his other doctrines, he must not advocate the above story more than any other, or else he would be going beyond what he said was possible to surmise about noumena. It is just unclear, then, how Kant means his readers to take the hints that led to the above story—maybe as a scratching of the unavoidable itch with which asking questions about the noumenal world leave us.

Actually, and more consistently, Kant does not think there is much, if anything, to say about the origin of evil in humans, which is what one would expect if said origin is indeed located in the human being as a noumenon. Before concluding this, however, Kant takes an interesting diversion to answer the question of how it is that moral evil spreads and propagates throughout times and people. I say it is an interesting diversion because it is not a question I believe we, as modern people, and for whatever reason, are tempted to ask. We take it for granted that the only evil imputable to an agent is his or her own—not anyone else's. Kant, then, may have been responding to a theme in the Christian and Jewish Scriptures, where sins of previous generations are said to justify retribution towards later generations. As Robert Adams notes in the introduction to one edition of the Religion, certain traditional doctrines of original sin state that human beings have actually inherited guilt from their first ancestors (Adam and Eve) (xii). Kant argues, to the contrary, that "every evil action must be so considered, whenever we seek its rational origin, as if the human being had fallen into it directly from the state of innocence" (6:41).

Moreover, Kant takes a somewhat metaphorical reading of the Scriptures that support the story he has been telling of the origin of human evil, where the temporal sequence of events of the Fall actually correspond to his ontological "sequence" (from the cause of free will to the effect that is moral evil considered morally). He explicitly links up his view with the Scriptural one by saying, "The mode of representation which the Scriptures use to depict the origin of evil, as having a beginning in human nature, well agrees with the foregoing…" (6:41). Of course, there is no corresponding beginning in time.

The passage in the section which immediately follows the one just under discussion adds a bit of confusion. Kant says, "…a prior innate propensity to transgression is presupposed in us but not in the first human being, in whom rather innocence is presupposed with respect to time" (6:42). Given all that has gone before, it is rather unclear why Kant would allow for any difference in the moral status (or propensities, or predispositions) of the first human beings. He seems

to think the difference is warranted as a result of the observation that "the first human being…is represented with full control of the use of his reason from the beginning" (6:43). It appears that the only way we can countenance this is to assume Kant is still in a sort of metaphorical mode (evidenced by the "represented" in the last quote). Even so, one might ask what purpose it serves him to represent the first human beings as innocent in the beginning. Perhaps the only answer is that he was trying to remain somewhat within the Biblical idiom, and this proved a useful vehicle in which to portray the necessity that evil be the result of a free act.

But at last we come to the conclusion of the section. After having said not much more than he was expected to say (namely, that the origin of human evil, in a rational sense, is not to be located in time, in some causal chain we can think about), Kant fesses up to his rather disappointing conclusion: "We must not however seek an origin in time of a moral character for which we are to be held accountable, however unavoidable this might be if we want to explain the contingent existence of this character" (6:43). In other words, the fact that we understand that the propensity to moral evil is not a necessary feature of human beings as a race (or individually) raises the question of why it is nonetheless a feature of human beings as we see them. Kant's admonishment might seem merely to ignore the question. However, it is evident that his hands are well and truly tied, given how he put moral evil squarely at the feet of human free will. (It should be noted, of course, that the objection applies equally well to anyone who wants humans to be responsible for their evil actions, and so is not a particularly good objection to Kant's views.)

Still, we should never have expected more out of a Kantian theory of evil which is consistent with Kant's views about phenomena and noumena. This observation, coupled with the various explanations and readings given throughout the course of this paper, shows that Kant's views about the origin of evil are at least consistent. In other words, his arguments can be read so that they are indeed valid. Whether they are sound or not, of course, will depend on many other factors, and in particular on features of the noumenal world (including, importantly, its existence). Unfortunately (or maybe

fortunately, from the perspective of criticism), consistency is the best that Kant can (and maybe wants to) obtain for his theory, since by the theory itself we cannot gain any knowledge about the noumenal realm.

Notes

Wood, Allen, and di Giovanni, George, eds & trs. Kant, Immanuel: *Religion within the Boundaries of Mere Reason And Other Writings.* Cambridge, UK: Cambridge University Press, 1998.

Wood, Allen. "Kant and the Problem of Human Nature". [Available Online]. http://www.stanford.edu/~allenw/recentpapers.htm. Also in: B. Jacobs and P. Kain, eds. *Essays on Kant's Anthropology.* New York: Cambridge University Press, 2003.

About The Author

Jonathan Lipps is a recent graduate of Stanford University with his B.A. and M.A. in Philosophy.

© 2007 ouragora. All rights reserved.