

# ORIGINAL SIN, RADICAL EVIL AND MORAL IDENTITY

Philip L. Quinn

In this paper, I consider Kant's theory of radical evil as an attempt to rationalize the Christian doctrine of original sin. As I read Kant, he is committed to the view that some ways of understanding the doctrine of original sin express in an inadequate fashion an important truth about human moral character. His project in Book One of *Religion Within the Limits of Reason Alone* is to formulate that truth within the bounds of reason drawn by his mature practical philosophy. The project is executed with Kant's characteristic subtlety and originality. It is astonishing to discover how much of what seems gripping in the intuitions which underlie the doctrine of original sin can be accommodated within the limits of Kantian reason. Yet, as I shall argue, the project ultimately fails, for its execution leads Kant into inconsistency. This inconsistency, however, is itself instructive because it illuminates starkly a severe tension within certain Christian views of moral character, a tension Kant was too sensitive to ignore but unable to resolve.

Put rather simply, the doctrine of original sin says that all humans, except Adam, Eve, Mary and Jesus, are born bearing a burden of sin and, hence, guilt. But it seems impossible that anyone should literally be sinful and guilty antecedent to having acted wrongly. And so, on the face of it, it looks as though the doctrine of original sin is incoherent. Lazy intellects may be tempted to suggest that the doctrine should nonetheless be believed as a mystery of faith and to leave it at that. But the solution of the lazy intellect is too quick to be philosophically acceptable. If some version of the doctrine is to be worthy of our belief, it must at least be internally consistent, that is, the propositions formulating it must be mutually consistent, and externally consistent, that is, it must also be consistent with everything else we know. Of course, this does not mean that a Christian has an epistemic duty to abandon belief in any particular version of the doctrine merely because it looks as though it is incoherent. A version of the doctrine may be both internally and externally consistent despite appearances to the contrary. Arguments which purport to establish internal inconsistency may be unsound, and claims of external inconsistency may be based on mistaken views of what else we do know. But the methodological moral for Christian philosophers is that, at minimum, it is in the long run incumbent upon them to defeat arguments intended to show that the doctrine is internally or externally inconsistent if they wish to hold that in the long run



believing the doctrine is not irrational.

Christian philosophers might hope to achieve a bit more. To secure some version of the doctrine of original sin against arguments for its internal or external inconsistency it would be sufficient to show that the version in question is internally and externally consistent. Or it might be argued that a doctrine very much like traditional versions of the doctrine of original sin is both internally and externally consistent, and for that reason that doctrine might be recommended for belief as a rational replacement for the traditional doctrine. Kant's aim, I think, is to achieve this much and a bit more. His doctrine of radical evil is intended as a rational replacement for traditional versions of the doctrine of original sin. He recommends it for belief on the grounds that it is internally consistent, that it is externally consistent with everything we can know within the framework of his mature practical philosophy, and that, in addition, it is supported by inductive evidence to the extent such support is possible within that framework. I propose to subject these claims to critical scrutiny.

The paper is divided into three sections. In the first section, using one traditional model for understanding the doctrine of original sin, I construct a formal argument to show that no one now bears the burden of original sin. The purpose of this argument is to bring into sharp focus the problem confronting Kant. In the second section, I show how Kant deploys the resources of his practical philosophy to modify the traditional model in a way that allows him consistently to reject one of the premisses of that argument and yet to attribute radical evil to every human. In the third section, I argue that the theory of moral character Kant uses to formulate his doctrine of radical evil is inconsistent with certain further claims he makes about the possibility of changing one's moral character. I conclude by suggesting that the underlying presuppositions about moral character which lead to this inconsistency reflect some deep but problematic features of Christian thinking about human nature and the moral life.

### *The Problem Articulated*

One rather crude model for understanding original sin that has appeared in the history of Christian thought involves transmission by physical inheritance. On this model, Adam and Eve physically transmitted sinfulness and, hence, guilt to their progeny, and this process has continued, leaving unaffected only Mary and Jesus, and will continue through all generations of offspring of our first human ancestors. Kant dismisses this inherited disease model as "inept" (p. 35)<sup>1</sup>, but it is much worse than that. From four plausible principles concerning original sin so conceived, it can be shown that no one now bears the burden of original sin. The principles are these:

- (A) If anyone now bears the burden of original sin, then he is such that

there is some fault with respect to which he is sinful and which he has physically inherited.

(B) If anyone is such that there is any fault with respect to which he is sinful, then he is guilty of bringing about that fault.

(C) If anyone is guilty of bringing about any fault, then he is accountable for that fault.

(D) If anyone is such that there is any fault which he has physically inherited, then he is not accountable for that fault.

And the argument, which can be cast in the form of a *reductio*, is this:

(1) Suppose someone now bears the burden of original sin. (Hypothesis for *reductio*)

(2) He is such that there is some fault with respect to which he is sinful and which he has physically inherited. (From 1 and A)

(3) He is such that he is sinful with respect to that fault. (From 2).

(4) He is guilty of bringing about that fault. (From 3 and B)

(5) He is accountable for that fault. (From 4 and C)

(6) He is such that he has physically inherited that fault. (From 2)

(7) He is not accountable for that fault. (From 6 and D)

(8) He is accountable for that fault and he is not accountable for that fault. (From 5 and 7)

(9) Hence, no one now bears the burden of original sin. (From 1 through 8)

Since the argument is formally valid, at least one of the four principles, (A)-(D), on which it is based must be rejected if its conclusion is to be reasonably denied. Various strategies which involve rejecting the inherited disease model have been tried out. I shall mention in passing two of them.

As is well known to Christian philosophers, Jonathan Edwards tried to defend the doctrine of original sin by rejecting the claim that there is real personal identity through time. On the Edwardian view, persons are no more than aggregates of successive, momentary person-stages. Since it is just for God to treat later Adam-stages as one with the Adam-stage which ate the forbidden fruit, and hence as sinful, even though they did not eat the fruit, it is no less just for God to treat the stages of Adam's descendents as one with Adam's stages, and so also as sinful, even if they do not literally sin or physically inherit any taint from any of Adam's stages. Thus Edwards would reject principle (A). He would say that the person-stages presently existing do bear the burden of original sin, or at least that it is just and reasonable for God by convention to treat them as though they do, despite the fact that some of them are such that there is no fault with respect to which they are literally sinful, guilty and accountable, and even if there is no fault which they have physically inherited from Adam. Bizarre though it seems, the Edwardian theory of

persons appears to be internally consistent. I have argued elsewhere that it is externally inconsistent, or at least that we are justified in taking it to be so, and I shall not repeat those arguments here.<sup>2</sup> I will, however, venture the opinion that the Edwardian theory is not a particularly plausible candidate for the status of rational replacement for the inherited disease model of original sin.

Kant mentions in passing another strategy which has a certain theological interest. According to this Platonistic view, Adam's descendents do not really begin to exist at some time in the interval from conception to birth but existed contemporaneously with Adam. And so when Adam first sinned, "we ourselves participated (although now unconscious of having done so)..." (p. 36). Someone of this opinion would also reject principle (A) and the inherited disease model. Such a person would say we bear the burden of original sin because Adam's sin itself is a fault with respect to which we are sinful, guilty and accountable on account of our free participation in it, and so of course we do not acquire this fault by physical inheritance from Adam. As far as I can tell, this view is internally consistent, and I doubt that any of us now is in a position to demonstrate that it is externally inconsistent. But like the Edwardian view, it strikes me as lacking promise as a rational replacement for the inherited disease model.

Kant too may be read as rejecting principle (A) and the inherited disease model. But his alternative strikes me, initially at least, as a promising and plausible replacement for that model. And so I now turn to the task of explicating the Kantian alternative.

### *Radical Evil in Human Nature*

In discussing the contents of Book One of the *Religion*, I shall focus on those points which are salient for appreciating how Kant's doctrine of radical evil constitutes a rationalization of the doctrine of original sin. Emphasis will be placed on Kant's views of human nature, willing and moral character. In order not to interrupt the flow of my exegetical argument, I shall relegate discussion of my disagreements with some prominent commentators to the footnotes.

According to Kant, human nature contains an original predisposition (*Anlage*) to good that can be functionally analyzed into three constituents, themselves predispositions, which are elements in the fixed character and destiny (*Bestimmung*) of humankind. First, there is the predisposition to animality, which Kant subsumes under the general concept of "physical and purely *mechanical* self-love, wherein no reason is demanded" (p. 22). This predisposition consists of instinctive drives for such things as self-preservation, propagation by means of sexuality, and the society of one's own kind. It is something we have in common with other animal species in virtue of a shared animal nature. Second, there is the predisposition to humanity, which Kant subsumes under the general concept of self-love which is

not merely instinctual but also rational. It is based on practical reason, Kant tells us, “but a reason thereby subservient to other incentives” (p. 23), namely, the incentives rooted in our sensuous nature. Hence, the kind of practical reason involved in the predisposition to humanity is instrumental or means-end reasoning of the sort Hume had in mind when he declared reason to be the slave of the passions. What is distinctive about Kant’s view is that he holds that reason can also be “practical of itself” (p. 23), that is, dictate moral laws unconditionally, and so he claims a third predisposition for human nature. This is the predisposition to personality, which we have in virtue of being both rational and accountable for our actions. Kant describes it as “the capacity for respect for the moral law as in itself a sufficient incentive of the will” (pp. 22-23). It is, of course, an indication of the importance of the moral realm in Kant’s mature philosophy that he includes what he takes to be distinctive of us as rational moral agents among the characteristic marks of human nature.

To say that these predispositions are original is, for Kant, to say that “they are bound up with the possibility of human nature” (p. 23). In other words, an individual could not lack any of them and still be a normal, mature human; they are essential attributes of normal, mature humans. It is for this reason that Kant calls them elements in the fixed character and destiny of man. To say they are predispositions to good is, for Kant, to say that they “are not only *good* in negative fashion (in that they do not contradict the moral law); they are also predispositions *toward good* (they enjoin observance of the law)” (p. 23). Of course no one is morally good simply in virtue of possessing them, for no one is accountable for having them. But since they predispose us toward moral good, whence then comes moral evil? If we have essentially the capacity for respect for the moral law to serve by itself as a sufficient incentive of the will, why is that capacity not exercised in each morally significant action we perform? Why does respect for the moral law sometimes fail to serve as a sufficient incentive for the will?

According to Kant, the ultimate answers to these questions are to be given in terms of a propensity to moral evil. Because for Kant moral evil must be tied up with the exercise of a free will<sup>3</sup> and a free will can be morally evil only in virtue of the maxims it adopts, a propensity to moral evil must be “the subjective ground of the possibility of the deviation of the maxims from the moral law” (p. 24). As the ground in the agent for the possibility of the deviation of the maxims of particular actions from the moral law, the propensity to moral evil must be in the agent antecedent to the adoption by the agent of the maxims of particular actions. Hence it must be represented as innate in the agent. And yet because Kant is convinced that a propensity to moral evil is itself something morally evil, it must also result from an exercise of the agent’s will for which the agent is accountable. And so Kant needs to claim that a propensity (*Hang*) can be distinguished from a predisposition (*Anlage*) by the fact that “although it can indeed be innate, it *ought* not to be rep-

resented merely thus; for it can also be regarded as having been *acquired* (if it is good), or *brought by man upon himself* (if it is evil)" (p. 24).

But how is this possible? If a characteristic is innate in a person, then it would seem that person is no more accountable for it than for anything else he or she got in the natural lottery. Innate attributes, even if they are accidental rather than essential to their possessors, seem to be things which are themselves neither morally good nor morally evil. If a characteristic of a person is morally evil, then it would seem that person acquired it through some particular free act or acts rather than possessing it antecedent to all such particular acts. Characteristics which are morally evil and for which one is accountable seem not to be things which could be innate in an agent. Outright inconsistency seems to be lurking just around the corner.

Kant thinks he has a way to avoid such inconsistency. If moral evil is to be attributed to a person or to a person's character and not just to particular actions the person freely performs, then we must attribute to the person not only particular evil maxims for his or her particular morally evil free actions at various times but also "an underlying common ground, *itself a maxim*, of all particular morally-evil maxims" (p. 16, my italics). Kant tells us we may think of such a maxim as a rule the will makes for the use of its freedom which "applies universally to the whole use of freedom" (p. 20). Such a maxim must be supreme in the sense that, because its adoption is the work of a free will, we can no more find a further causal explanation for it "than we can assign a cause for any fundamental attribute belonging to our nature" (p. 27). Because the adoption of a supreme maxim is an exercise of freedom, it is an act for which the agent is accountable. The propensity to moral evil is therefore itself a morally evil act in this sense and is also the ground of all morally evil particular actions. But it is, Kant goes on to say, "intelligible action, cognizable by means of pure reason alone, apart from every temporal condition" (pp. 26-27). And just because it can be understood to be this peculiar sort of atemporal free act which is nevertheless the common ground underlying the whole use in time of an agent's freedom, it can be represented as always in the agent serving as a ground for all his or her particular actions at particular times. For this reason, it is not misleading to represent a propensity to moral evil both as innate in the sense that it is present in the agent independent of and therefore antecedent to all exercises of freedom in time and as inextirpable in the sense that it cannot be eradicated by means of exercises of freedom in time.

Thus the picture of the propensity to moral evil which emerges from Kant's discussion is this. The agent's will freely but independent of all temporal conditions adopts an evil supreme maxim. The evil supreme maxim is a rule or general policy which covers every exercise of the agent's freedom in time and thereby provides a ground in the agent for all his or her particular morally evil actions at particular times. This general policy does not causally necessitate particular actions at particular times; particular actions at particular times too must be exercises of freedom

if they are to be morally good or evil. This is why its adoption only grounds the *possibility* of the deviation of the maxims of particular actions from the moral law and why it establishes only a *propensity* to moral evil in particular actions. Nevertheless, the propensity is itself morally evil because it springs from an exercise of freedom, and it can be represented as being present in the agent before, during and after all exercises of the agent's freedom in time in virtue of being understandable as being independent of all temporal conditions. It is not the sort of thing an agent could gain or lose in the course of time. However, it is what makes an agent's character, as opposed to his or her individual actions, morally evil, and so it provides the agent with moral identity. An agent has a morally evil character just in case he or she has a propensity to moral evil.

Since the propensity to moral evil is a product of freedom, it cannot be an essential element in human nature as is the predisposition to good. If moral evil is to be attributed to mankind as a species, it must be a contingent and accidental attribute of each member of the species. For this reason, the claim that man is evil 'by nature' is not to be justified by philosophical analysis of human nature. Its justification, Kant tells us, is rather "that from what we know of man through experience we cannot judge otherwise of him, or, that we may presuppose evil to be subjectively necessary to every man, even to the best" (p. 27). As Kant reads the historical evidence, it provides plenty of inductive support for such a presupposition. Primitive peoples are barbaric and cruel, civilized peoples are vicious and depraved, and nations are no better than primitive individuals. Kant's pessimism puts him on the side of Hobbes in his assessment of the moral evil in the state of nature and on the side of Rousseau in his judgment of the moral evil in the civilized state. And so Kant takes himself to have good inductive support for attributing a morally evil propensity to moral evil to mankind universally.

But it seems very improbable that a propensity to moral evil should be both a product of freedom and universal among mankind. Because the adoption of an evil supreme maxim is an absolutely spontaneous exercise of the will, it is antecedently likely that some people would have freely adopted a morally good supreme maxim while others adopted a morally evil supreme maxim. Even if it is impossible to assign numerical values to the prior probabilities of the various alternatives, it seems clear enough that the prior probability of all humans choosing freely a morally evil supreme maxim must be quite low. Being well aware of this, Kant tells us that the freedom of the choice of a supreme maxim "will not tally with the universality of this evil *unless* the ultimate subjective ground of all maxims somehow or other is entwined with and, as it were, rooted in humanity itself" (pp. 27-28). So Kant sees quite clearly that the *universality* he attributes to the propensity to moral evil needs an explanation of some sort. Yet how can he root the propensity in human nature itself without compromising or qualifying his claim that it is a product of freedom for which people are accountable?

Kant considers and rejects two possible answers to this question. The propensity to moral evil cannot originate in the natural inclinations arising from our sensuous nature. In the first place, they are part of the predisposition to animality, which Kant consistently holds to be a predisposition to good; indeed, “they afford the occasion for what the moral disposition in its power can manifest, namely, virtue” (p. 30). In the second place, because “they are implanted in us” and “we are not their authors” (p. 30), we are not accountable for them, though we are accountable for the propensity to moral evil. Nor can the propensity to moral evil originate in a corruption of the morally legislative function of reason. A being in whom opposition to the moral law was an incentive to action would, Kant thinks, be a devilish being; such a being would lack the predisposition to personality, which is essential to normal, mature humans. In the case of man, the moral law “forces itself upon him irresistibly by virtue of his moral predisposition” (p. 31). A human simply cannot, as Kant sees it, flat out renounce the moral law in the manner of a rebel; no matter what a human being does, the moral law cannot cease to exert its influence as an incentive.<sup>4</sup>

According to Kant, “freedom of the will is of a wholly unique nature in that an incentive can determine the will to an action *only so far as the individual has incorporated it into his maxim* (has made it the general rule in accordance with which he will conduct himself); only thus can an incentive whatever it may be, co-exist with the absolute spontaneity of the will (i.e., freedom)” (p. 19). Incentives by themselves never causally determine the will to action; a will so determined would not be free. Incentives, so to speak, clamor for our attention and press us to act on them. But it is up to us to decide which incentives shall determine the will to action by choosing freely how incentives shall be incorporated into the maxims of our actions. In virtue of our natural predispositions, we are always confronted with incentives of two sorts. In virtue of the predisposition to personality, respect for the moral law is an incentive for us; in virtue of the predisposition to animality, the inclinations which arise from our sensuous nature are also incentives for us. If incentives of either sort were present alone and incentives of the other sort absent, the agent would adopt incentives of the sort present into his or her maxim and those incentives would thereby come to serve as the sole and sufficient determining ground of the agent’s will. But since incentives of both sorts will inevitably press us so long as we remain human, the will must cope with or respond to incentives of both sorts whenever it frames a maxim for action. Because incentives of the two sorts sometimes press for the same action but sometimes conflict, it will only be accidental if both sorts of incentives press the agent in the same direction. So if the will’s maxim is to have the form of a general or generalizable rule, it must subordinate incentives of the one sort to those of the other in order to cover cases where incentives conflict. A good maxim is one in which the incentives of inclination are subordinated to the incentive of respect for the moral law; an evil maxim is one

which lacks this subordination.

At the level of the choice of a supreme maxim which is to serve as a policy for the whole use of one's freedom, Kant will admit only two possibilities. Either the supreme maxim is good or it is evil; Kant's rigorism allows for no third alternative. Kant tells us that it is "of great consequence to ethics in general to avoid admitting, so long as it is possible, of anything morally intermediate, whether in actions (*adiophora*) or in human characters" (p. 18). This rigorism is intended to rule out the possibility of moral characters which are either (i) morally good in some respects and morally evil in others or (ii) neither morally good nor morally evil. So the good person's policy is to make pursuit of the satisfaction of his sensuous inclinations *always* subordinate to obedience to the moral law. Any other policy is the policy of an evil person, for it lacks the proper moral order of subordination of incentives. This implies that Kant is committed to holding that it would be evil even to incorporate the moral incentive and the incentives of inclination into one's maxim on an equal footing in cases where incentives of both sorts press toward the same action. And in discussing impurity, which he regards as one of three degrees in which the propensity to moral evil manifests itself, Kant does say that "although the maxim is indeed good in respect of its object (the intended observance of the law) and perhaps even strong enough for practice, it is not yet purely moral; that is, it has not, as it should have, adopted the law *alone* as its *all-sufficient* incentive" (p. 25). Any policy which is not purely moral because it does not make respect for the law its sufficient incentive and subordinate all other incentives to respect for the law must, on the rigoristic view, be an evil policy. And so when Kant tells us that "man (even the best) is evil only in that he reverses the moral order of the incentives when he adopts them into his maxim" (p. 31), we must, if we wish to attribute to Kant a consistent view, understand reversals of the moral order of the incentives broadly as *any* failure to preserve the proper subordination of the incentives of inclination to the moral incentive and not narrowly as merely those failures which consist in subordinating the moral incentive to the incentives of inclination. To subordinate the moral incentive to the incentives of inclination is to exhibit wickedness, which Kant regards as the third and worst degree in which the propensity to moral evil manifests itself and which he characterizes as "the propensity of the will to maxims which neglect the incentives springing from the moral law in favor of others which are not moral"<sup>5</sup> (p. 25).

A person's failure to preserve the proper and purely moral order of incentives when he or she incorporates them into a supreme maxim explains how the propensity to moral evil can be both rooted in human nature, in that it is made possible by the inevitable presence in humans of incentives of two kinds which can, and often do, come into conflict, and also a product of freedom, in that it consists of a free choice of a supreme maxim which does not always give obedience to the moral law pride of place over pursuit of the satisfaction of sensuous inclinations. And it

seems reasonable enough to suppose that such a policy is, implicitly at least, universal among mankind.

As I said above, a propensity to moral evil in an agent does not causally necessitate particular morally evil actions at particular times. Particular actions too are products of the freedom of absolute spontaneity. And so Kant can consistently maintain that it is possible for an agent to “overcome” (p. 32) the propensity to moral evil in that he or she can perform many good actions in the course of time despite the presence of the propensity to moral evil in him or her. What an agent cannot do is to extirpate or eradicate the propensity in the course of time because it is an act independent of all temporal conditions which is the ultimate ground in the agent of all maxims of particular actions at particular times.

Kant’s theory of a universal human propensity to moral evil which is itself morally evil seems to me a very good candidate for the job of rationalizing the Christian doctrine of original sin. I think Kant’s theory does a measure of justice to several important intuitions which underlie the doctrine of original sin. First, this propensity is a moral evil which is a product of freedom and for which the agent is accountable. Second, it is radical in the sense that it goes to the root of the exercise of human freedom and is, so to speak, entwined with human nature. Third, it is constitutive of the agent’s moral character and identity and, being universal among human agents, may be said to belong to humans ‘by nature’. And, fourth, because it is the result of an act independent of temporal conditions, it may be represented from the point of view of the temporal realm as present in the agent in such a way as to be antecedent to all particular actions at particular times and, hence, innate and inextirpable by the agent in the course of time. I know of no rational reconstruction of the doctrine of original sin that preserves more of these intuitions than Kant’s theory of radical evil does. Moreover, as far as I can tell, the theory is internally consistent or, more cautiously, is logically possible if it is logically possible for there to be human acts of will which are both absolutely spontaneous and independent of all temporal conditions. While I have no proof that such acts are possible, I also know of no proof that they are impossible. And although such acts, if they occurred, would admittedly have to be inexplicable and mysterious in some ways, I believe we can get some idea of how they might occur, if only by analogy, from the theological discussion of how spontaneous acts of will are possible for a timeless God.

#### *Radical Transformation of Moral Identity*

It would have been nice if Kant had stopped at this point in the exposition of his theory and said no more in Book One of the *Religion*. Alas, he did not. Kant also holds that human agents are capable of effecting a revolution whereby a morally evil supreme maxim is replaced by a morally good supreme maxim. Thereby the

person becomes, so to speak, a new man “by a kind of rebirth, as it were a new creation (John III, 5; compare also Genesis I, 2), and a change of heart” (p. 43). How is it possible for a person who begins as a morally evil person to become by his or her own powers a morally good person? That it is possible Kant has no doubt, for he holds that because we ought to do, we can do it, since nothing is demanded of us by morality which we cannot do. Kant thinks we can avoid problems if we say “that man is under the necessity of, and is therefore capable of, a revolution in his cast of mind, but only of a gradual reform in his sensuous nature (which places obstacles in the way of the former)” (p. 43). And he goes on to explain that “if a man reverses, by a single unchangeable decision, that highest ground of his maxims whereby he was an evil man (and thus puts on the new man), he is, so far as his principle and cast of mind are concerned, a subject susceptible of goodness, but only in continuous labor and growth is he a good man” (p. 43).<sup>6</sup> Such a reversal would, Kant thinks, restore to the predisposition to personality its original purity and power.

But Kant confronts a serious difficulty here and knows it, for he goes on to ask whether such a restoration through one’s own efforts is inconsistent with the assumption of an antecedently inherent morally evil propensity to moral evil. He answers: “Yes, to be sure, as far as the conceivability, i.e., our *insight* into the possibility, of such restoration is concerned” (p. 46). But he claims that the assumption in question is not really inconsistent with the possibility of such a restoration just because “when the moral law commands that we *ought* now to be better men, it follows inevitably that we must *be able* to be better men” (p. 46). Unfortunately, these remarks do not solve the problem. We not only lack insight into how it is possible for a human agent both to have a morally evil propensity to moral evil and to reverse the highest ground of his or her maxims by a single unchangeable decision. In addition, it can be demonstrated, I believe, that these two things are not compossible, and so we have positive insight into their joint impossibility.

More precisely, if Kant supposes that a moral revolution ever occurs, then he is committed to an inconsistent triad of claims. One is the thesis of radical evil itself, according to which every human by an act independent of all temporal conditions has adopted a morally evil supreme maxim that makes his or her moral character evil and that is inextirpable. The second is the thesis of rigorism, according to which a human’s moral character cannot be a mixture of good and evil, and so can be constituted by at most one supreme maxim, and cannot be neither good nor evil, and so must be constituted by at least one supreme maxim. And the third is the thesis of moral revolution, according to which some human by a single unchangeable decision reverses the highest ground of his or her maxims and adopts a morally good supreme maxim.

Now consider the following line of argument. Suppose some human adopts a morally good supreme maxim. Either a morally good supreme maxim coexists in

that human with a morally evil supreme maxim, or it does not. If it does, then that human has two supreme maxims, one good and the other evil, and his or her moral character is a mixture of good and evil, which contradicts the thesis of rigorism. If it does not, then either that human has not adopted a morally evil supreme maxim or that human had at one time adopted a morally evil supreme maxim and at some later time given it up and replaced it with a morally good supreme maxim. But the assumption that that human has not adopted a morally evil supreme maxim contradicts the part of the thesis of radical evil which asserts that every human has adopted a morally evil supreme maxim. And the supposition that that human had at one time adopted a morally evil supreme maxim and at some later time given it up contradicts the part of the thesis of radical evil which asserts that a morally evil supreme maxim, because it is the product of an act independent of all temporal conditions, is inextirpable once adopted. Hence, no human adopts a morally good supreme maxim.

It is important to be clear about just what I take this argument to establish. It does not show that no human can adopt a morally good supreme maxim; in other words, it does not show that there is no possible world in which some human does freely adopt a morally good supreme maxim. Indeed, any argument which demonstrated this would show too much, for it would establish that all humans adopt a morally evil supreme maxim of necessity. What the argument does show is that there is no possible world in which (i) the thesis of rigorism is true, (ii) every human adopts a morally evil supreme maxim, and (iii) some human adopts a morally good supreme maxim. But this suffices to establish that it is not possible for the thesis of rigorism to be true, for every human to adopt a morally evil supreme maxim, and for some human to adopt a morally good supreme maxim. And this in turn is enough to show that it is not possible, given the thesis of rigorism, for a moral revolution, that is, a change or reversal of supreme maxims, to occur in any single human person. If a person adopts a morally evil supreme maxim, then any person who adopts a morally good supreme maxim is literally a different person and so, literally and not just 'as it were', a new man and a new creation. A difference in moral identity this deep requires a difference in personal identity if we are to remain within the bounds of Kant's theory. Or, at any rate, so I have argued.

And if we ask, in a Kantian spirit, whether we ought to effect a moral revolution in our characters by a reversal of supreme maxims from evil to good, then I think we should conclude that we ought not because we cannot. It is, perhaps, a tribute to Kant's humaneness that he wishes to hold out to us the prospect of a revolution in moral character. Unfortunately, the logic of his own theory forbids it. All that theory allows is incessant struggle against the propensity to moral evil by means of particular good actions. Thus, Kant's attempt to rationalize the Christian doctrine of original sin must be judged a failure because his theory of radical evil is inconsistent with the conjunction of two other claims equally central to his mature practi-

cal philosophy, the thesis of rigorism and the thesis of moral revolution.

Could another philosophical theory succeed where Kant's theory has come to grief? The question is obviously a difficult one, and I can offer no definitive answer. However, I should like to conclude by making some tentative and rather speculative remarks about why this appears to me unlikely.

One approach would be to abandon the attempt to rationalize the doctrine of original sin within the categories of Kant's philosophy and to try to reconstruct that doctrine with the resources of some other conceptual framework. To take just one example from post-Kantian thought, one might construe the assertion that almost all human persons bear the burden of original sin as claiming that almost all human persons are in a state of alienation from God. However, a mere terminological shift obviously will not solve all the problems Kant grappled with. With respect to any human who is alleged to be in this alienated state, we may ask whether his or her being in such a state is a product of the exercise of his or her freedom or not. If it is not, then it is a mere misfortune and not something he or she is guilty of or accountable for, and so it is at best misleading to assimilate it to the category of sin. If it is, then one must work out some piece of theory similar to the disease model of physical transmission or Kant's model of a free act independent of all temporal conditions to explain how it can be, apparently, innate in the moral character. In short, the problems Kant confronted and tried to solve are real, and they cannot be dissolved or made to disappear merely by a switch of conceptual frameworks.

In fact, I believe these problems arise from a very deep tension in typically Christian views of the human situation. Let me try to express that tension in a stark and simple fashion. Two presuppositions are that human characters as well as individual actions are proper subjects of moral evaluation and that human characters are products of the exercise of freedom to the extent that they are properly so evaluated. On the one hand, there is the intuition that almost all human characters are innately and ineradicably stained in a way that is properly subject to a negative moral evaluation because it is somehow self-incurred. This somber assessment of the human situation can strike one as quite realistic when examining one's own inner life or the historical record in certain frames of mind. On the other hand, there is the intuition that only a human character which was spotless and so properly subject to a wholly positive moral evaluation could find favor with, and be justified in the eyes of, a morally perfect God. This demand for purity and perfection of moral character is just what one would expect from a God who is himself morally perfect in every respect. If these two intuitions are coupled with the hope that nonetheless some of us may be justified in the eyes of God in a way which is not capricious but is in accord with his essential perfect justice, then it may seem as if the moral characters experience teaches us we have must be completely different from the moral characters we would have to have if the hope were to have any foundation. And so the hope may seem sensible only if miraculous discontinuities

in moral character, breaks so sharp they would appear destructive of moral and perhaps even personal identity, are possible. In my opinion, Kant accepted the two presuppositions, found the two intuitions compelling, and tried to resolve the tension between them in order to make room for the hope. I consider Kant's attempt a splendid and instructive failure because it teaches by example how difficult the project of effecting a philosophical reconciliation between the two intuitions will be if it is carried out with rigor and attention to detail.

Confronting this tension, many theologians will attempt a resolution by appeal to the miracle of divine grace and mercy. Whether such an appeal is available to a philosopher, within the bounds of Kantian reason or outside those limits, and, if so, whether it can succeed must be topics reserved for another paper.<sup>7</sup>

*Brown University*

#### NOTES

1. Immanuel Kant, *Religion Within the Limits of Reason Alone*, translated by Theodore M. Greene and Hoyt H. Hudson with introductory essays by Theodore M. Greene and John R. Silber (Harper & Row: New York, 1960). All quotations from this work will be identified by parenthetical page references.
2. Philip L. Quinn, "Some Problems About Resurrection," *Religious Studies* 14, pp. 343-59, and "Divine Conservation, Continuous Creation, and Human Action," *The Existence and Nature of God*, edited by Alfred J. Freddoso (University of Notre Dame Press: Notre Dame, 1983) pp. 55-81.
3. I use the word 'will' to refer to what Kant calls *Willkür*, the power of choosing among alternatives. In the *Religion*, Kant attributes to *Willkür* libertarian freedom of a very strong sort, namely, absolute spontaneity. It is not my aim in this paper to argue for or against this libertarian position.
4. John R. Silber considers Kant in error on this point because he thinks that "man's free power to reject the law in defiance is an ineradicable fact of human experience" (John R. Silber, "The Ethical Significance of Kant's *Religion*," *Religion Within the Limits of Reason Alone* (Harper & Row: New York, 1960) p. cxxix). If by 'rejecting the law in defiance' Silber means eradicating the capacity for respect for the moral law to serve as a sufficient incentive of the will, then Kant would deny that a human can reject the law in defiance. Any being that could reject the law in this sense would not have essentially the predisposition to personality, which is an essential attribute of humans, and so would simply not be human. But it is not at all obvious from experience that humans can reject the law in this sense. And as Allen W. Wood has pointed out (Allen W. Wood, *Kant's Moral Religion* (Cornell University Press: Ithaca, 1970) pp. 212-13), the evidence Silber offers in support of his view is not very impressive. The fact that evil men in fact and fiction such as Hitler, Napoleon and Ahab have been particularly energetic and single-minded in their evil-doing hardly suffices to show that they are utterly without conscience to the extent of lacking even the *capacity* for respect for the law to serve as a sufficient incentive of the will. Such men as Hitler may seldom, if ever, have exercised that capacity, but this would not show they did not possess it.
5. Silber holds (*op. cit.*, p. cxxvi) that an evil disposition is manifested in wickedness but not in impurity. As I read Kant, both impurity and wickedness are manifestations of an evil disposition. In favor of

my interpretation, I can cite Kant himself. He tells us that the will's capacity not to adopt the moral law into its maxim arises from the propensity to moral evil, and presumably he refers to the will's capacity not to adopt the moral law into its maxim in the right way, namely, as its superordinate and sufficient incentive. And then he explicitly asserts that "in this capacity for evil there can be distinguished three distinct degrees" (p. 24) and goes on to enumerate frailty, impurity and wickedness. If one thinks, as I admit Kant's actual words sometimes suggest, that the evil alternative to a good supreme maxim, which subordinates the incentives of inclination to the moral incentive, is a maxim which involves subordination in the reverse order, then Silber's position will seem attractive because Kant says that wickedness is that degree of the capacity of evil which "reverses the ethical order (of priority) among the incentives of a free will" (p. 25). But if one thinks, as I do, that Kant is better interpreted as wishing to claim that any supreme maxim which fails to subordinate the incentives of inclination to the moral incentive is evil, then it is easy to understand how impurity, as well as wickedness, can be a degree within the capacity for evil that arises from the propensity to moral evil.

6. Allen W. Wood emphasizes that "we should not confuse this 'change of heart' with an actual 'change' of any kind in time" (Wood, *op. cit.*, p. 229.). And he goes on to say this: "Regarded as the supersensible ground for a life of moral progress, a good disposition is a kind of 'revolution', eternally counteracting and overcoming the innate propensity to evil through constant and morally motivated limitation and discipline of those inclinations through which this propensity manifests itself. But even the best disposition cannot *extirpate* the propensity to evil, and end man's self-incurred liability to transgression of the moral law" (*ibid.*, pp. 230-31). However, Wood does not discuss, and apparently does not see, the serious problem Kant's view will encounter when it comes time to explain how it is possible for a good supreme maxim to coexist with an inextirpable evil supreme maxim in one and the same person. In the next four paragraphs of the main text, I argue that Kant's theory does not have the resources to allow him to construct an acceptable solution to this problem.

7. Many of the ideas in this paper originated in a seminar on Kant's philosophy of religion I taught at Brown in 1981 and at Notre Dame in 1982. I thank the students in those seminars for helpful discussion. David Chalfin and Dayna Macy, whose MA theses grew out of papers they wrote for the seminar at Brown, were particularly valuable sources of intellectual stimulation.