Choosing Wickedness: Moral Evil in Kant's Religion

Changtze Chia
Purdue University

Follow this and additional works at: https://docs.lib.purdue.edu/open_access_dissertations

Recommended Citation
https://docs.lib.purdue.edu/open_access_dissertations/1703

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Jacqueline Mariña, Chair
Department of Philosophy
Dr. Daniel Smith
Department of Philosophy
Dr. Christopher Yeomans
Department of Philosophy
Dr. Christopher Insole
Department of Theology and Religion, University of Durham

Approved by:
Dr. Rod Bertolet
Head of the Graduate Program
ACKNOWLEDGMENTS

I would like to express my gratitude to:

Professor Jacqueline Mariña. She directed this dissertation from beginning to end. The half-baked ideas that I started with were developed into their present state under her patient guidance. She kept me on the straight and narrow, curbing my tendencies to take my dissertation in odd (and utter indefensible) directions. This dissertation would not exist without her. I am forever indebted to Professor Mariña for the second chance she gave me.

Professor Daniel Smith, Professor Christopher Yeomans, Professor Christopher Insole. They have been remarkably helpful throughout my writing of this dissertation. Their comments on what I had written were incisive and insightful. They each contributed greatly to improve my dissertation for the better. I am grateful to each and every one of them for their assistance, and honored by their willingness to sit on my dissertation committee.

My parents. For being beside me each and every step of the way. I would have given up long ago without their love and support.
# TABLE OF CONTENTS

LIST OF ABBREVIATIONS ........................................................................................................... vi

ABSTRACT ........................................................................................................................................ vii

INTRODUCTION ................................................................................................................................. 1

The Problem of Radical Evil ........................................................................................................... 2

Investigating Radical Evil ............................................................................................................... 5

CHAPTER 1. FREEDOM IN KANT’S PRACTICAL PHILOSOPHY .................................................. 14

Introduction ....................................................................................................................................... 14

Theoretical Freedom ......................................................................................................................... 15

Ought-Implies-Can ........................................................................................................................... 21

The Imputation Problem ................................................................................................................. 26

Spontaneity ....................................................................................................................................... 29

Autonomy and Heteronomy ............................................................................................................. 39

The Reciprocity Thesis ..................................................................................................................... 48

Conclusion ....................................................................................................................................... 54

CHAPTER 2. TWO CONCEPTIONS OF RADICAL EVIL ................................................................. 56

Introduction ....................................................................................................................................... 56

Radical Evil as Fundamental Maxim .............................................................................................. 58

The Anthropological Evidence for Radical Evil ............................................................................. 70

Radical Evil as Innate Disposition ................................................................................................. 80

Frailty and Depravity ....................................................................................................................... 88

Conclusion ....................................................................................................................................... 95

CHAPTER 3. THE NOUMENAL ORIGINS OF RADICAL EVIL ..................................................... 97

Introduction ....................................................................................................................................... 97

Noumenal Choice ............................................................................................................................. 99

Fallibility and Infallibility ............................................................................................................... 104

Vice and Virtue ................................................................................................................................ 111

Frailty and Embodiment ............................................................................................................... 120

Depravity and Self-Conceit .......................................................................................................... 126

Conclusion ....................................................................................................................................... 135
CHAPTER 4. INTERPRETATIONS OF RADICAL EVIL ............................................. 137

Introduction .................................................................................................................. 137

Chris Firestone and Nathan Jacobs ............................................................................. 138

Seiriol Morgan ............................................................................................................. 145

Henry Allison ............................................................................................................... 154

Allen Wood ................................................................................................................ 163

Conclusion .................................................................................................................. 176

CONCLUSION ............................................................................................................ 179

BIBLIOGRAPHY ........................................................................................................ 192
LIST OF ABBREVIATIONS

Page references to Kant’s works adhere to the standard practice of providing the volume and page number in *Kants gesammelte Schriften*, edited by the Königliche Preussische Akademie der Wissenschaften (Berlin: Walter de Gryter, 1902ff). The sole exception to this method occurs in references to Kant’s *Critique of Pure Reason*, in accordance with a different academic convention. Every page reference to Kant’s *Critique of Pure Reason* provides the edition (“A” for the 1st Edition and “B” for the 2nd Edition) and then the page number (e.g. A537/B565).

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Title</th>
<th>Translator(s)</th>
<th>Publisher</th>
</tr>
</thead>
<tbody>
<tr>
<td>VA</td>
<td><em>Anthropology from a Pragmatic Point of View</em></td>
<td>Robert Louden</td>
<td>Cambridge University Press, 2006</td>
</tr>
<tr>
<td>VE</td>
<td><em>Lectures on Ethics</em></td>
<td>Peter Heath and J. B. Schneewind</td>
<td>Cambridge University Press, 1997</td>
</tr>
<tr>
<td>VpR</td>
<td><em>Lectures on Philosophical Theology</em></td>
<td>Allen Wood and Gertrude Clark</td>
<td>Cornell University Press, 1978</td>
</tr>
<tr>
<td>MS</td>
<td><em>The Metaphysics of Morals</em></td>
<td>Mary Gregor</td>
<td>Cambridge University Press, 1996</td>
</tr>
<tr>
<td>Rel</td>
<td><em>Religion Within the Boundaries of Mere Reason</em></td>
<td>Allen Wood and George di Giovanni</td>
<td>Cambridge University Press, 1998</td>
</tr>
<tr>
<td>Th1</td>
<td><em>Theoretical Philosophy After 1781</em></td>
<td>Gary Hatfield, Michael Friedman, Henry Allison, and Peter Heath</td>
<td>Cambridge University Press, 2002</td>
</tr>
<tr>
<td>Th2</td>
<td><em>Theoretical Philosophy, 1755-1770</em></td>
<td>David Walford and Ralf Meerbote</td>
<td>Cambridge University Press, 1992</td>
</tr>
</tbody>
</table>
This dissertation concerns Kant’s doctrine of radical evil. The doctrine consists in the following two claims: (1) The propensity to evil is a universal propensity, found in every human being. It is therefore an innate disposition. (2) Every human being is morally responsible for the propensity to evil found in him. It is therefore a disposition that is freely chosen. The poor reception Kant’s doctrine of radical evil has received over the centuries can be traced to his insistence on these two claims. The doctrine has been censured for two reasons: (1) Kant appeals to empirical evidence in his argument that the propensity to evil is a universal propensity. But an empirical generalization cannot prove that the propensity to evil is universal. Interpreters of Kant have therefore tried to correct this huge lapse by providing us with Kant’s missing proof for the universality of radical evil. (2) Kant argues that the propensity to evil is both universal and freely chosen. But these are not two predicates that really go together. It is objected that the determinism necessary to make the disposition universal is logically incompatible with the demands of freedom, or that it is ridiculously improbable to suggest that every human being freely made the exact same (evil) choice. Interpreters of Kant have therefore tried to reconcile the two concepts by coming up with a plausible account of how every human being could have freely chosen evil. This dissertation will deal with both problems: (1) This dissertation argues that the problem of the missing proof is a non-problem. Kant’s argument for radical evil has been incorrectly interpreted as an empirical generalization. It is not. The examples of human evil doing that Kant gives are not meant to prove that human beings are universally evil. Kant is using them as counterexamples against the theory that human beings are universally good. (2) This dissertation argues that the universality of radical evil can be reconciled with the free choice of radical evil. The human being makes the noumenal choice of a fundamental maxim, which thereby determines his moral orientation. If he chooses to wholeheartedly embrace the moral law, he would have made the noumenal choice of a good fundamental maxim. Anything less than such a wholehearted and unconditional embrace of the
moral law represents a noumenal choice for the evil fundamental maxim. For he has thereby made
his obedience of the moral law conditioned upon something else. It is then argued that for a human
being to so wholeheartedly embrace the moral law is not a logical impossibility, but it is, however,
a real impossibility. In these two responses, this dissertation presents solutions to concerns that
surround Kant’s doctrine of radical evil; concerns that are collectively labeled as the “problem of
radical evil”.
INTRODUCTION

The initial publication of the Religion within the Boundaries of Mere Reason in 1792 was greatly anticipated by the Enlightenment milieu of the day. It proved to be a bitter disappointment. Kant’s contemporaries reviled the work, seeing the condemnation of human nature as evil as a betrayal of the Enlightenment values of freedom and human progress. Their reactions to the doctrine of radical evil were captured in Goethe’s letter to Herder in 1793, in which he had written: “Kant required a long lifetime to purify his philosophical mantle of many impurities and prejudices. And now he has wantonly tainted it with the shameful stain of radical evil, in order that Christians too might be attracted to kiss its hem” (Fackenheim 1954, p. 340). But the Religion fared little better in the eyes of the Prussian censors. Kant’s insistence that the human being was himself responsible for the evil of his nature led him to dismiss the Biblical account of original sin with the scathing words: “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 quae non fecimus ipsi, vix ex nostra puto” (6:40). The doctrine of radical evil turned out to be an anathema to both sides. The Enlightenment progressives despised it because they considered the universality of radical evil an affront to the freedom of choice that every human being has in determining his own moral character. The conservative Prussian bureaucrats censured it because they considered the adoption of radical evil through the free power of choice a danger to the universality of the evil of human nature. By ascribing to

1 This is translated as: “Race and ancestors, and those things which we did not make ourselves, I scarcely consider as our own” (Ovid, Metamorphoses, XIII: 140-141).
radical evil the characteristics of freedom and universality, Kant pleased precisely no one, but instead created a conceptual tension in the very idea of radical evil that has persisted to this day as the problem of radical evil.

**The Problem of Radical Evil**

What, precisely, is the issue with radical evil being both universal and free? Here is one formulation of the problem of radical evil: Radical evil can be universally present in every human being, or it can be freely adopted by every human being that has this propensity, but it cannot be both universally present in every human being and freely adopted by every human being with this propensity. This would be a logical inconsistency. Quinn gives us an example of the reasoning at work here in the following passage:

> If a characteristic is innate to 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 the things which could be innate in an agent. Outright inconsistency seems to be lurking just around the corner (Quinn 1984, p. 193).

The issue here has to do with Kant’s claim that the evil in human beings is innate. It is because this evil is innate that radical evil can be said to be universal, that is, present in every human being:

> “Whenever we therefore say, “The human being is by nature good,” or, “He is by nature evil,” this only means that he holds within himself a first ground (to us inscrutable) for the adoption of good or evil (unlawful) maxims, and that he holds this ground qua human, universally – in such a way, therefore, that by his maxims he expresses at the same time the character of his species” (6:21–22).

But, as Quinn notes, if a characteristic is innate to a person, then he cannot be held accountable for having that characteristic. And, this line of reasoning continues, if radical evil is innate to the
human being, then it cannot be imputed to any human being that has it, because it cannot have an origin in freedom – it must have its origins somewhere other than freedom, such as in the laws of nature.

This formulation of the problem of radical evil therefore reprises the age-old conflict between freedom and thoroughgoing causal determinism described in the Third Antinomy. It is represented as a dilemma. The first horn of the dilemma is this: Insofar as radical evil is free, it can only be produced through a special kind of causality, namely a faculty of absolutely beginning a state, and hence also a series of its consequences (A445/B473). And the second horn of the dilemma is this: Insofar as radical evil is innate, it can only be produced through the laws of nature, where everything that happens presupposes a previous state, upon which it follows without exception according to a rule (A444/B472). It is therefore contradictory to hold both to be true, because radical evil cannot have an origin in freedom if it is innate, and radical evil cannot be innate if it is to have an origin in freedom. This is the problem with radical evil that Michalson hits upon when he points out the logical conflict between freedom and innateness (universality) in the following passage: “Moreover, because Kant thinks that the propensity to evil is universal (“this propensity can be considered as belonging universally to mankind…”), he is finally led to the seemingly paradoxical judgment that evil is both freely elected and “innate”, a view that would appear to robe either “freedom” or “innateness” of its point” (Michalson 1990, p. 46). If understood in this way, there is no special problem of radical evil. The problem of radical evil is merely a different way of formulating the problem of the Third Antinomy.

But there is another formulation of the problem of radical evil: Radical evil can be universally present in every human being, and it can be freely adopted by every human being that has this propensity, but it seems highly unlikely that radical evil is both universally present in every
human being and freely adopted by every human being that has this propensity. This would be improbable, perhaps even to the point of being a statistical impossibility. Here again we have Quinn to give us an example of the reasoning at work here in the following passage:

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 human beings choosing freely a morally evil supreme maxim must be quite low (Quinn 1984, p. 194).

The issue here is the remarkable coincidence that Kant is asking us to accept. Suppose we were talking about a coin-toss. If we flip a coin, we can expect it to come up heads about half the time and tails about half the time. Imagine, then, if we flipped the coin a thousand times in a row, and the coin came up heads every time. Even though it is mathematically possible for the coin to come up heads a thousand times in a row, the likelihood of such an event happening is infinitesimally small, and it would be wholly unreasonable for either of us to assume that there is no trickery afoot. The universal presence of radical evil in every human being confronts us with the same problem. Here we are being asked to believe that every human being spontaneously exercises his freedom in the exact same way – to choose the evil fundamental maxim. It would be no more reasonable for us to believe Kant’s assertion that all these choices were made freely and without external compulsion, than for you to believe my assertion that these thousand coin tosses that came up heads were the result of chance. If understood in this way, there is a special problem of radical evil. For Kant must be able to explain precisely what it is about the moral psychology of human beings that compels them, but does not determine them, to freely choose the evil fundamental maxim without exception. And this explanation has a high hurdle to overcome: It must persuade
us that it is reasonable to believe that billions upon billions of human beings spontaneously made
the same free choice.

Investigating Radical Evil

In this essay, the problem of radical evil will be treated primarily as a problem of moral
psychology. That is, we will treat the problem of radical evil as having to do with the psychological
processes that lead a human being to freely choose radical evil, and determine whether this process
can plausibly be thought to occur in every human being in order to produce a universal propensity
to evil. Let us do this by breaking down the problem that lies before us into several parts: First, we
must ask ourselves how the doctrine of radical evil relates to the rest of Kant’s practico-
philosophical writings. For these writings are the source of the moral concepts on the basis of
which Kant posited the doctrine of radical evil. Second, we explain what is this something that
Kant calls radical evil. From earlier, we know that he ascribes to radical evil the characteristics of
freedom and universality. But this cannot be the entire story, since it does not tell us what radical
evil is, that is, we have not specified what it means for radical evil to be described as a propensity,
or as a fundamental maxim. Third, a solution to the problem of radical evil as a problem of moral
psychology must be provided. That is, we must in some way make it reasonable to believe that
every human being to have ever lived freely chose the evil that has taken root in his heart.
Therefore, this essay will be structured in accordance with the following four questions:

1. What is the place of freedom in Kant’s practical philosophy?
2. What is radical evil?
3. What makes the universal propensity to evil a real possibility?
4. What solutions have other authors given for the problem of radical evil?
In Chapter 1, we will be considering the place of freedom in Kant’s practical philosophy. This chapter is structured around the Kantian distinction between the Wille and the Willkür. For Kant, the Wille is the legislative aspect of the will, and the Willkür is the executive aspect of the will. It is the Wille that provides us with the moral law, and it is the Willkür that then makes the choice whether to obey or to disobey this law. Freedom, then, means something very different for the Wille than it does for the Willkür, because the Wille is free insofar as it is an autonomous lawgiver, whereas the Willkür is free insofar as it is a spontaneous decider. The main discussions of this chapter will consist in the following: First, the problem of imputation. Put briefly, the issue is this: Kant asserts that we are morally responsible for the evil that we do. This can only be the case if we freely chose to perform those evils. Freedom is a necessary condition for those acts to be imputed to us, otherwise we would not be able to take responsibility for them. But Kant also asserts that freedom and morality reciprocally entail each other. If so, this means that the concept of the moral law entails that of freedom, and vice versa. And since evil acts are by definition contrary to the moral law, it would seem that we must deny that they are free. This leaves us with the intolerable conclusion that our evil acts, being unfree, cannot be imputed to us for the purposes of moral responsibility. This sets up the discussion of freedom as it relates to the Wille and the Willkür, which comes second. We begin by considering freedom in the sense of spontaneity. This distinction pertains to freedom as it relates to the Willkür. When the will is free in the sense of spontaneity, it is free to perform an action and it is also free to refrain from performing said action. This means, in the case of evil acts, a will that is spontaneously free has the capacity to subordinate the moral incentives to incentives that are unmoral, thereby making the maxim of the deed (and the deed itself) evil, and the same goes for good acts, mutatis mutandis. It is because we performed the evil act when we could have refrained that makes the deed blameworthy. Moral evil is therefore
imputable to us because we have a free Willkür. This may be contrasted with freedom in the sense of autonomy. This distinction pertains to freedom as it relates to the Wille. The freedom of the Wille is autonomy, and it has to do with the will’s own self-legislation. A will that is unfree in this sense does not self-legislate. It is a Wille that is heteronomous. And this means that the will is not a law unto itself, but instead obtains its laws from sources outside itself. It is freedom in the sense of autonomy that implies and is reciprocally implied by the moral law. Third, we will turn to consider the Reciprocity Thesis more closely. This is Kant’s idea that freedom and morality reciprocally entail each other. Here we will argue that Kant’s idea is that autonomy and morality reciprocally entail each other, and not that spontaneity and morality reciprocally entail each other. Thus, the solution to the problem of imputation will take place along these lines: Evil acts are unfree as they relate to the Wille, because they are the product of heteronomy rather than autonomy. But evil acts are free as they relate to the Willkür, because we could have done otherwise, and this makes us responsible for the evil acts that we perform.

In Chapter 2, our discussion shifts to the topic of radical evil. It is concerned with the following cluster of problems: What is a propensity to evil? What does it mean for a fundamental maxim to be evil? Are they one and the same thing? We need to ask these questions because Kant describes radical evil in many ways. In some places, he describes radical evil as the propensity to evil in human beings. This way of describing radical evil makes it an innate disposition, albeit an innate disposition that may be represented as either acquired or brought by the human being upon himself (6:29). In other places, however, radical evil is described as the evil fundamental maxim that is found in every human being. And this way of describing radical evil makes it the subjective first ground of our maxims, that is, the ground of all our maxims that deviate from the moral law (6:22). First, we turn to discuss radical evil in terms of an evil fundamental maxim. The argument
that leads Kant to posit the existence of such a maxim turns on the way maxims are connected to each other in his practical philosophy. For Kant, a maxim can only be grounded by another maxim. And once it is understood that every one of our maxims are grounded in some way or the other, this leads us to conclude that there must be a hierarchy of maxims. **Second**, we consider the anthropological evidence for radical evil that Kant provides in the *Religion*. In this section, it is argued that Kant’s argument for the universality of the propensity to evil has been misunderstood. The point of these examples he provides is not to serve as an inductive proof that human beings are by nature evil. Rather, they are examples used to falsify the thesis, popular in Kant’s day, that human beings are by nature good. **Third**, we explain how the anthropological evidence from before can be used to prove that the propensity to evil is *universal*. Briefly put, the anthropological evidence of widespread evil suffices to prove that evil is *possible* for every human being. And a condition for this possibility is the *actual* presence of a dispositional property in every human being that would activate if the conditions were just right, in accordance to which the human being would commit evil. **Fourth**, we consider the gradations of evil that Kant introduces in the *Religion*, specifically the evils of frailty and depravity. Insofar as the evil of frailty seems to involve no maxim, it can only be produced through a subjective first ground for the possibility of an inclination, that is, through the propensity to evil. The evil of depravity, by contrast, does involve a maxim. Indeed, it cannot take place except through a maxim. As such, it must have been produced through the subjective first ground for the adoption of a maxim, that is, through an evil fundamental maxim. **Thus**, it is difficult to see how the propensity to evil and the evil fundamental maxim can ever be one and the same thing, that is, radical evil. And even though there is an argument that establishes a universal propensity to evil, it seems to be an equivocation to say, on this basis, that *radical evil* is universal.
In Chapter 3, our topic remains fixed on the nature of radical evil. Here, however, our interests are slightly different, because we are asking: What are the conditions that make the universal propensity to evil a real possibility? Prima facie, the intelligible world is the ground of the sensible world and its laws, from which it follows that the intelligible character of an individual is the ground for the empirical individual and the laws of his behavior (his empirical character). If the propensity to evil is universal, then it follows that every human being chooses not to incorporate the moral law in the construction of his fundamental maxim. This solution to the problem of radical evil leaves precisely no one satisfied. It provides a solution that is merely formal—a trivial metaphysical corollary of the metaphysics of transcendental idealism—and not a solution in terms of Kant’s moral psychology. In the end, we still would not know what it means for Kant to say that every human being makes the choice of evil in the constitution of his fundamental maxim. By delving into the real possibility of radical evil as an evil fundamental maxim, we can begin to flesh out the moral psychology involved in the making of this choice. First, we ask how it is possible to adopt a fundamental maxim that is evil when we, as atemporal first causes, are making our decisions in the noumenal realm? This seems impossible. Moral evil, Kant tells us, is only possible when we subordinate the moral incentive in favor of other incentives that are nonmoral. These other incentives that are nonmoral are not available to us in the noumenal realm. Ipso facto, we cannot adopt fundamental maxims that are evil. And yet there is a universal propensity to evil, as demonstrated in the previous chapter, and this would be impossible if we had not all selected a fundamental maxim that was evil. Second, assuming that we have no choice but to incorporate the moral incentive into our fundamental maxim, then either we are not morally responsible for our choice of fundamental maxim, or we have a choice in how we incorporate the moral incentive into our fundamental maxim. In this section, we will be opting for the latter option.
The difference in the way we incorporate the moral incentive into our fundamental maxim determines whether our will is fallible or infallible. If the moral law is something we have steadfastly incorporated into our hearts, then the will would be infallible, because there are no circumstances under which it can be led astray. But if the moral law is compromised in us, then the will would be fallible, and it will go astray the moment an opportunity arises for it to. Third, the fallibility or infallibility of a will is a matter of its moral strength. Kant calls this moral strength virtue (6:380). Every individual has incorporated the moral incentive into his fundamental maxim, and therefore has moral strength to some degree or the other. But only in an infallible will is this moral strength perfected (maximally present), whereas in a fallible will, the moral strength is lacking, and the degree to which the fallible will lacks moral strength is called moral weakness. This moral weakness is only a potential for going astray, because there is nothing in the noumenal realm that can actually lead the will astray. Afterwards, we will conclude this chapter with a discussion of frailty and depravity. These gradations of evil arise from necessary phenomenal structures that determine the nature of the human being. These are original predispositions that are present in every human being. The predisposition of animality is the source of the human being’s mechanical self-love. Every human being is embodied, and insofar as he is embodied, every human being is subject to sensible desires and inclinations. The self-love that aims at the satisfaction of a human being’s animal desires is mechanical self-love. The moral weakness of a human being, insofar as it is unable to withstand these sensible desires and inclinations when they oppose the moral disposition in him, gives rise to the evil of frailty. Likewise, the predisposition of humanity is the source of the human being’s comparative self-love. Out of (comparative) self-love, every human being is determined to gain worth in the opinion of others (6:27) But this self-love gives way to self-conceit, because this desire to gain equal worth in the opinion of others gradually
becomes “an unjust desire to acquire superiority for oneself over others” (6:27). This transition from self-love to self-conceit is depravity, and the moral fault for it lies with the human being, for he could have prevented this slide into depravity had he not been morally weak in character. **In conclusion,** by demonstrating how the choice of a fundamental maxim interacts with the necessary phenomenal structures that are a part of human nature, we explain how the evil that is grafted onto processes that occur in every human being is able to produce a universal propensity to evil. In so doing, we provide an account of radical evil in terms of Kant’s moral psychology.

In **Chapter 4,** we will review a limited selection of the extensive literature on radical evil. We will be asking the following question: How have other authors understood (and provided their own solutions to) the problem of radical evil? Each of the proposals we will be examining attempt to make sense of Kant’s assertion that radical evil is both universal to the human species and freely chosen by every human being. **First,** we consider the proposal put forth by Firestone and Jacobs in their book *In Defense of Kant's Religion.* There they appeal to the distinction between primary substance and secondary substance from Aristotle’s metaphysics, where the individual and the species it belongs to are both considered substances by virtue of their being subjects of predication. The idea here is that, although it would be hard to imagine multitudes of individuals all making the same free choice to adopt radical evil, this is not what happens. Instead, it is the human species, as a (secondary) substance, that makes this free choice to adopt radical evil, and the propensity to evil is universal by fact of our participation in the human species. **Second,** we turn to the proposal put forth by Morgan in his article “The Missing Formal Proof of Humanity's Radical Evil in Kant's Religion”. Morgan interprets the choice between radical evil and the original good in terms of a choice between two different conceptions of freedom. Negative freedom is the autonomy of a will that has freed itself from determination by external causes. Such a will would, according to Kant,
legislate to itself the universal law of reason. For Kant, negative freedom is supposed to lead to positive freedom. But the desire for negative freedom can also be corrupted into a desire for untrammeled freedom, that is, a freedom that is beholden to no law, not even the moral law. And it is the desire for negative freedom in this sense that leads us to embrace radical evil. Third is Allison's argument proving the universality of radical evil. Allison contrasts the propensity to evil with the propensity to good. Per ethical rigorism, the human being has to have one or the other. By unpacking the propensity to good in terms of Kant's concept of the holy will, Allison interprets the propensity to good as a spontaneous obedience to the dictates of the moral law. The finite will, being incapable of ignoring completely the demands of inclination, can therefore never be called holy. And since the holy will and the propensity to good are one and the same, it must be concluded that the propensity to evil must be present in every finite will. Fourth, we finally arrive at the anthropological account of radical evil. The version we will be dealing with here is Wood's. According to Wood, every human being tends towards unsociable sociability, that is, they make comparisons and compete amongst themselves for self-worth. The human being finds himself in a society of human beings, and he instinctively works to gain worth in the opinions of others, not allowing anyone else superiority over himself for the sake of self-protection. This gradually gives way to an unjust desire to acquire superiority over others for oneself, and it is this that Wood identifies as the propensity to evil. But each of these attempts at a solution to the problem of radical evil contain a flaw: they all trace the genesis of radical evil to a sufficient condition in the nature of humanity. This methodology is inconsistent with the way Kant characterizes radical evil. He writes: “He is evil by nature” simply means that being evil applies to him considered in his species; not that this quality may be inferred from the concept of his species (from the concept of a human being in general, for then the quality would be necessary)...” (6:32). In providing radical evil with
a sufficient condition, these authors turn radical evil into a quality that may be inferred from the concept of his species. They make radical evil a universal certainty by sacrificing its freedom.
CHAPTER 1. FREEDOM IN KANT’S PRACTICAL PHILOSOPHY

It matters not how strait the gate,
   How charged with punishments the scroll,
I am the master of my fate,
   I am the captain of my soul.

William Ernst Henley, *Invictus*.

**Introduction**

The Critical period of Kant’s philosophical career began with the publication of the *Critique of Pure Reason* in 1781. In this work, Kant hoped to “deny knowledge in order to make room for faith” (Bxxx). For he had been troubled by the advancement of dogmatic metaphysics that purported to explain the world so exhaustively and completely that there was no longer space to assume the existence of God, freedom, and the immortality of the soul, even for the practical use of reason. This practical use of reason can be seen in Kant’s insistence that moral evil, as opposed to physical evil, has to be the product of our freedom: “… A physical propensity (one based on sensory inducements) to whatever use of freedom, be it for good or evil, is a contradiction. Hence a propensity to evil can only attach to the moral faculty of choice. Nothing is, however, morally (i.e. imputably) evil but that which is our own deed” (6:31). As this passage makes clear, whether or not we are morally responsible for our actions hinges on whether it was a matter of our choice. For if we were not free in performing those actions, then they cannot be imputed to us, and we bear no moral responsibility for either the action, or the consequences that arise from it. But it is also through freedom that the moral law is given to us. Freedom, after all, is the *ratio essendi* of the moral law, and the moral law is the *ratio cognoscendi* of freedom, as Kant had declared in the famous passage from the Preface of the *Critique of Practical Reason* (5:4n). Freedom and the moral law reciprocally imply each other: “For, had not the moral law already been distinctly
thought in our reason, we should never consider ourselves justified in assuming such a thing as freedom (even though it is not self-contradictory). But were there no freedom, the moral law would not be encountered at all in ourselves” (5:4n). These are two distinct concepts that both attach to Kant’s idea of freedom, and each of these concepts need to be clearly and distinctly explained, as they will be here, otherwise we would arrive at no precise understanding of why Kant was so insistent that radical evil had to be freely chosen by every human being.

**Theoretical Freedom**

In the First Critique, Kant’s discussion of freedom is featured most prominently in the Third Antinomy. The Third Antinomy introduces the opposition between fatalism and freedom. Fatalism is described in the Third Antinomy in terms of a thoroughgoing causal determinism: “Assume there is no causality than that in accordance with the laws of nature: then everything that happens presupposes a previous state, upon which it follows without exception according to a rule” (A444/B472). The fatalist position can be contrasted with that of freedom, which is described in the Third Antinomy as the idea of an uncaused cause: “Suppose there were a freedom in the transcendental sense, as a special kind of causality in accordance with which the occurrences of the world could follow, namely a faculty of absolutely beginning a state, and hence also a series of its consequences; then not only will a series begin absolutely through this spontaneity, but the determination of this spontaneity itself to produce the series, i.e., its causality, will begin absolutely, so that nothing precedes it through which this occurring action is determined according to constant laws” (A445/B473). In this passage, freedom is described as the capacity to absolutely begin a causal series, that is, both the state and the series of consequences that are supposed to follow from it. The Thesis of the Third Antinomy makes the case for freedom through a reduction ad absurdum of fatalism. And the Antithesis of the Third Antinomy makes the case for fatalism
by doing the same to freedom. This makes the antinomial conflict between fatalism and freedom intractable: Each side of the disjunction has a knock-down argument against the other.

Kant’s solution to the antinomial opposition between fatalism and freedom was to argue that there is a way for both reductios to be correct: “If, on the other hand, appearances do not count for any more than they are in fact, namely not for things in themselves but only for mere representations connected in accordance with empirical laws, then they themselves must have grounds that are not appearances. Such an intelligible cause, however, will not be determined in its causality by appearances, even though its effects appear and so can be determined through other appearances” (A537/B565). In this way, the argument of the Thesis and the argument of the Antithesis can be reconciled with each other. Insofar as the category of causality is to be limited in application to only those objects intuited in time and space, the appearances are determined without exception in accordance with the laws of nature. Kant is therefore well within his rights to assert that every state that happens is preceded by a state from which it follows according to a rule, because the only causality present in the world of appearances is this causality of nature. But this does not force him to give up on the existence of a causality of freedom that is distinct from the causality of nature. For there can be a world beyond the sensible world, and the intelligible world would be governed by laws of its own, distinct from the laws of nature that govern the sensible world. If so, then there could well be an entire world that is governed, not by the laws of nature, but by the law of morality, and the determinations that are the result of this law’s operation would be brought about through a special kind of causality that Kant calls the causality of freedom. Thus, it is in no way inconsistent that an appearance has a sensible cause, in which it is determined by another appearance in accordance with the laws of nature, and also an intelligible cause, in which
it is determined from the side of the things themselves by a power that is able to absolutely begin a state, and hence also a series of its consequences.

But this solution is a negative one. It attempts to show that fatalism and freedom are not inconsistent, given the metaphysics of transcendental idealism. This is a far cry from showing that there actually is such a faculty of absolutely beginning a state. In his writings circa the First Critique, Kant made attempts to provide positive arguments for the existence of freedom that were disjunctive in nature, in which he focused on disproving (or making untenable) the fatalist position detailed in the Third Antinomy. Hence, in his review of Schulz's *Sittenlehre* of 1783, Kant writes: “Although he would not himself admit it, he has assumed in the depths of his soul that understanding is able to determine his judgment in accordance with objective grounds that are always valid and is not subject to the mechanism of merely subjectively determining causes, which could subsequently change; hence he always admits freedom to think, without which there is no reason” (8:14). The thrust of Kant’s argument in the Review rests on the practical incoherence of the fatalism that Schulz advocated. Fatalism here refers to Schulz’s denial of the reality of freedom, a position against which Kant writes: “… The most confirmed fatalist, who is a fatalist as long as he gives himself up to mere speculation, must still, as soon as he has to do with wisdom and duty, always act as if he were free, and this idea also actually produces the deeds that accords with it and can alone produce it. It is hard to cease altogether to be human” (8:13). In other words, the problem with the fatalist's denial of freedom is that it is self-undermining.

To see why, consider what the fatalist position entails: there is no freedom, and all the actions of the will are determined by the laws of nature, all appearances to the contrary. This means that the fatalist must regard his actions to be, one and all, the product of natural necessitation. Even so, the fatalist is prepared to give arguments in defense of his fatalistic commitments, and assumes
that he will be able to convince his skeptical colleagues of their rationality. This attitude does not square with the fatalist's denial of freedom. For fatalism considers the fatalist's belief of fatalism to itself be the product of the necessitation of causal laws. And similarly, however brilliantly the fatalist makes his case, his colleague's positions for or against fatalism cannot be the result of his convincing arguments, but must also be the product of necessary causal laws. In other words, the fatalist's actions belie his commitment to fatalism, because what the fatalist does is indeed the opposite of what he should do, if his commitment to fatalism was genuine. Thus, according to Kant's approach in the Review, the fatalist's denial of freedom is contradicted by his own actions every time he uses reason to try to persuade himself and others to adopt fatalism.

This line of reasoning resurfaces in the next major work of Kant's corpus, the Groundwork of the Metaphysics of Morals (1785). There, the argument Kant provides takes on the following form:

Now, one cannot possibly think of a reason that would consciously receive direction from any other quarter with respect to its judgments, since the subject would then attribute the determination of his judgment not to his reason but to an impulse. Reason must regard itself as the author of its principles independently of alien influences; consequently, as practical reason or as the will of a rational being it must be regarded as free, that is, the will of such a being cannot be a will of his own except under the idea of freedom, and such a will must in a practical respect thus be attributed to every rational being (4:448).

Like before, Kant's argument in the Groundwork sets up a contrast between freedom and fatalism. In this argument, Kant is not focusing on reason in its practical use, but on reason in its speculative use. This means that he is focusing on the ability of reason more generally to make judgments. For a judgment to be rationally justified, reason must have determined that it had objective validity and was universally valid, and hence necessary, for everyone. This determination, however, would be meaningless if it was not made on the basis of universal rational principles, but on the basis of sensible impulse, "since the subject would then attribute the determination of his judgment not to
his reason but to an impulse” (4:448). The fatalist that denies freedom believes that everything is “determined to activity by the influence of alien causes” (4:446). Because this includes the power of judgment, this means that the fatalist must consider judgment to be an activity that is determined by the influence of alien causes. But a judgment that is determined by the influences of alien causes (the impulses of sensibility) would no longer be rationally justified in any meaningful sense. Even if this judgment has necessary universal validity, this necessary universal validity does not come because everyone has examined the judgment on the basis of universal rational principles and arrived at universal agreement. Rather, it comes about because everyone was determined by sensible impulse and the relevant laws of nature to arrive at universal agreement. In this sense, Kant’s claim that “reason must regard itself as the author of its principles independently of alien influences” echoes the claim he made in the Review that the fatalist (secretly) regards himself as the author of his principles independently of alien influences.² Hence: “Although he would not himself admit it, he has assumed in the depths of his soul that understanding is able to determine his judgment in accordance with objective grounds that are always valid and is not subject to the mechanism of merely subjectively determining causes, which could subsequently change; hence he always admits freedom to think, without which there is no reason” (8:14).

The argument Kant deploys against fatalism in the Review and the Groundwork focuses on showing that reason in its speculative use must be considered free. If it is not, then the claim to

² In his analysis of this reasoning, Wood notes that it is structurally similar to Kant’s argument that humanity is an end in itself. For this reason, the two arguments are problematic in the same way: “Neither is a deductively valid argument for its conclusion. All the premises of each argument, and even the proposition consisting in their conjunction, are quite consistent with the falsity of the conclusion. It may be impossible for you to do what you must represent yourself as doing when you act, judge, or even think, except by presupposing that you are free. But that is consistent with your not being able to do these things, and therefore with your not being free. In the same way, it may be impossible for you to set ends according to reason without presupposing that rational nature is an end in itself, but that is also entirely consistent with rational nature’s not having this value, and even with nothing’s having such a value” (Wood 2008, p. 133). Kant’s argument, as Wood observes, is not meant to place the falsity of its conclusion outside the realm of possibility, but to place it “out of reach for anyone to rationally deny the conclusion, or even to decline to assent to it” (Wood 2008, p. 133).
objectivity made in every judgment can no longer be epistemically justified. It would turn out that something is judged to be true, not based on rational principles, but because we were causally determined by sensible impulse to judge it to be true. The fatalist position that considers everything to be determined in their activity by natural causes is therefore incoherent in the sense that the fatalist can never practice what he preaches. But then Kant extends his argument to cover all rational beings, which he did not do in the Review: “Reason must regard itself as the author of its principles independently of alien influences; consequently, as practical reason or as the will of a rational being it must be regarded of itself as free, that is, the will of such a being cannot be a will of its own except under the idea of freedom, and such a will must in a practical respect be attributed to every rational being” (4:448). It is at this point that the argument shifts from a “first person” standpoint to a “third person” standpoint (Wood 2008, p. 133). For Kant is no longer arguing that a rational being cannot coherently deny his own freedom, but that a rational being cannot coherently deny freedom to any rational being. The reasoning for this shift is not clear, but Wood has suggested the following thought: “If I am going to reason with others, even about any theoretical matter, I must presuppose in them the same capacity to govern their judgments by rational norms that I must presuppose in myself” (Wood 2008, p. 133).³ This idea goes back to our earlier point about what it means for a judgment to be rationally justified. In principle, every rational being can assent to the judgment because it was made on the basis of universal rational

³ Wood cites a passage from the First Critique to justify this line of reasoning: “Reason must subject itself to critique in all its undertakings, and cannot restrict the freedom of critique through any prohibition without damaging itself and drawing upon itself a disadvantageous suspicion. Now there is nothing so important because of its utility, nothing so holy, that it may be exempted from this searching review and inspection, which knows no respect for persons. The very existence of reason depends on this freedom, which has no dictatorial authority, but whose claim is never anything more than the agreement of free citizens, each of whom must be able to express his reservations, indeed even his veto, without holding back” (A738/B766). The interaction with others as rational beings for the purpose of coming to a judgment’s necessary universal validity “is a necessary condition for the very existence of reason” (Wood 2008, p. 133). And this interaction would be impossible unless we attribute freedom of spontaneity to their reason, otherwise any necessary universal validity we arrive at would be hollow and meaningless.
principles – it has necessary universal validity. This condition is not met when we regard only our own reason as free, because the necessary universal validity of the judgment would be a sham if other rational beings besides us made their determination not based on reason but because of an impulse. Therefore, if we think of judgment in terms of its necessary universal validity, then it becomes evident that the “first person” standpoint of the Review is inadequate, and Kant’s argument in the Groundwork had to move beyond it to adopt the “third person” standpoint.4

Ought-Implies-Can

The principle that ought-implies-can asserts that, if there is a particular action that we ought to perform, then we have the ability to perform that action. Kant has traditionally been credited as the originator of this principle. Examples of this principle at work can be seen in the Religion: “... But if a human being is corrupt in the very ground of his maxims, how can he possibly bring about this revolution by his own forces and become a good human being on his own? Yet duty commands that he be good, and duty commands nothing but what we can do” (6:47). And also in the following passage from the Metaphysics of Morals: “Impulses of nature, accordingly, involve obstacles

4 Following Ameriks, we consider Kant to have abandoned his attempts to provide a theoretical proof of the existence of freedom after the Groundwork. According to Ameriks: “... The clear insistence that the representation of spontaneity in our thought does not yield knowledge of the self in itself gives Kant for the first time a consistently Critical theory of self-knowledge and simultaneously undercuts the first and third steps of the Foundations’s deduction: now our rationality cannot get us to a free noumenal self, and whatever noumenal side the self is allowed to have, the restrictions on ascribing any personal character to it eliminate even the possibility of arguing from freedom to morality on the basis of what is (supposedly theoretically) known to be one’s ‘proper self’. Thus, by the time of his latest work, Kant had totally abandoned the distinctive points of the last section of the Foundations” (Ameriks 2003, p. 183). This sets up the practical proof of the existence of freedom that becomes prominent in the Second Critique, in which “freedom is to be argued for only after morality is accepted” (Ameriks 2003, p. 183). From this time onwards, the starting-point for Kant’s proofs of the existence of freedom no longer rested on the spontaneity of our thought, but on a Fact of Reason: “Consciousness of this fundamental law may be called a fact of reason because one cannot reason it out from antecedent data of reason, for example, from consciousness of freedom (since this is not antecedently given to us) and because it instead forces itself upon us of itself as a synthetic a priori proposition that is not based on any intuition, either pure or empirical...” (5:31). And having determined the existence of the moral law in us, Kant is able to establish the existence of freedom as a postulate of pure practical reason, because “had not the moral law already been distinctly thought in our reason, we would never consider ourselves justified in assuming such a thing as freedom (even though it is not self-contradictory)” (5:4n).
within the human being's mind to his fulfillment of duty and (sometimes powerful) forces opposing it, which he must judge that he is capable of resisting and conquering by reason not at some time in the future but at once (the moment he thinks of duty) he must judge that he can do what the law tells him unconditionally that he ought to do” (6:380). These passages serve a twofold purpose: Not only are they proof that this principle has its origins in Kant's practical philosophy, but they also demonstrate the relevance of the principle to the core issues of moral imputability and practical freedom.

Now whence comes this “ought” in the principle that ought-implies-can? For Kant, ought-statements are imperatives: “All imperatives are expressed by an ought and indicate by this the relation of an objective law of reason to a will that is by its subjective constitution is not necessarily determined by it (a necessitation)” (4:413). This is because the finite will is constituted in such a way that its free power of choice cannot be determined to action through any incentive except so far as the human being has incorporated it into his maxim (6:24). Therefore, when the objective law of reason represents an end to such a will as something that is (morally) good, it is given an incentive to adopt a particular course of action. But merely having an incentive does not by itself lead to action – the incentive has to be incorporated into a maxim in order for an action to happen. And if the incentive given to us by the objective law of reason is not incorporated into a maxim, then no action determined by the thought of duty would result. By contrast, a will that is necessarily in accord with the moral law is not subject to imperatives: “A perfectly good will would, therefore, equally stand under objective laws (of the good), but it could not on this account be represented as necessitated to actions in conformity with law since of itself, by its subjective constitution, it can be determined only through the representation of the good” (4:414). Therefore, the objective law of reason prescribes actions to holy wills and finite wills alike, but it is only in the latter that these
prescriptions are mere incentives to action and not sufficient determining grounds of action. And therefore it is only in a finite will that these prescriptions take the form of imperatives, that is, as oughts.\footnote{Another issue would be how we are to understand the “can” of the principle that ought-implies-can. Meyer writes in his homily on Ecclesiastes 8:4 (1894): “Whatever He bids you to do by his word, be sure that He will enable you to do it by his power. He works in us to will and to work of his good pleasure; that is, He never directs us in any path of obedience or service without furnishing a sufficient supply of grace. Does He bid you to renounce some evil habit? The power to renounce it awaits you. Claim it. Does he bid you to walk on the water? The power by which to walk only waits for you to claim it.” Kant himself seems to share this sentiment when he writes: “Granted that some supernatural cooperation is also needed to his becoming good or better, whether this cooperation only consist in the diminution of obstacles or be also a positive assistance, the human being must nonetheless make himself antecedently worthy of receiving it; and he must accept this help (which is no small matter), i.e. he must incorporate this positive increase of force into his maxim: in this way alone is it possible that the good be imputed to him, and that he be acknowledged a good human being” (6:44). Does the “can” of the principle that ought-implies-can refer to what we can do through our power alone, or does it refer to what we can do through the supernatural assistance provided by grace, if only we accept this help from God? If it is the former, then we are left with the implication that God’s divine commands are not identical with the imperatives of morality, because the principle holds for the moral law but not for God’s divine commands, contra Kant’s statement later in the Religion that true duties must be represented as at the same time divine commands (6:99). But if it is the latter, then we are left with the implication that there can be an entire class of actions that we ought to perform, even though they are miraculous in nature, and therefore completely outside the natural order.}

If freedom were to be denied, then there would be no sense in which what ought to be is something that can be. Practical freedom has traditionally been understood in terms of the principle of alternate possibilities. According to the principle of alternate possibilities, an action is only free if the agent that performed it could have done otherwise. Kant describes freedom in terms of the principle of alternate possibilities in the following passage from the Religion: “... According to freedom, the action, as well as its contrary, must be in the control of the subject at the moment of happening” (6:50n). The principle that ought-implies-can features prominently in Kant’s discussion of freedom. This is because Kant thinks that there is a divergence between what is and what ought to be. This is evident in the following passage from the First Critique: “... Reason does not give in to those grounds that are empirically given, and it does not follow the order of things as they are presented in intuition, but with complete spontaneity it makes its own order according to ideas, to which it fits the empirical conditions and according to which it even declares actions..."
to be necessary that have not occurred and perhaps will not occur" (A548/B576). Therefore, if there was no freedom, then there are actions that reason declares to be necessary that cannot occur given the order of things as they are presented in intuition. Insofar as the order of things as they are presented in intuition would then be the only possible order of things, it follows that the order of things that reason comes up with would have to be dismissed as mere fantasy. It is not an alternate possibility because it is a nomological impossibility. Freedom, as the power to make alternate possibilities actual, has the role of making what ought to be into what is. Without freedom, we would have to give up the principle that ought-implies-can, because we would have no way of reaching those alternate possibilities that are specified by the “ought” in the principle that ought-implies-can.

This would be to the detriment of moral imputability. Moral imputability is the idea that immoral actions can be imputed to, or blamed on, the agents that are responsible for them, and the same goes for moral actions, mutatis mutandis. It is evident that Kant subscribes to the idea of moral imputability, as we can see from the following passage: “... But this subjective ground must, in turn, itself always be a deed of freedom (for otherwise the use or abuse of the human being’s power of choice with respect to the moral law could not be imputed to him, nor could the good or evil in him be called “moral”)” (6:21). This passage shows that the human being is to be blamed when he abuses his power of choice with respect to the moral law and praised in the proper use of his power of choice with respect to the moral law. The principle that ought-implies-can features prominently in this argument for the moral imputability of our actions. For Kant, it is the moral law that prescribes to us what we ought or ought not do, and we are to be praised or blamed depending on whether we have performed our duty. And this means that, if the principle that ought-implies-can does not hold true, it would be possible for the moral law to prescribe to us a duty that
we do not have the ability to fulfill. Insofar as we did not fulfill our duty, it is imputed to us as an abuse of our power of choice with respect to the moral law, and the end result is that we would be blamed for something that is outside our control. But surely we cannot be justifiably blamed for something that is outside our control. Therefore, the moral imputability of our actions must presuppose the principle that ought-implies-can.

The denial of freedom would therefore result in a disconnect between the moral law and moral responsibility. For it entails that there are actions that we ought to perform, but do not have the ability to perform. The moral law dictates the actions that we ought or ought not to perform. When we neglect to perform these actions, we act contrariwise to the moral law, and we bear moral responsibility for our actions. Freedom determines what we can or cannot do. It is by the actions we perform that we are able to depart from the order of things as they are presented to us in intuition. If the principle that ought-implies-can is affirmed, then the moral law dictates the actions that we ought or ought not perform, and freedom allows us to depart from the order of things in intuition to perform these actions. Hence, by neglecting to perform these actions, we have failed to properly exercise our freedom. And when something ought to have happened (but did not), or ought not to have happened (but did), we are rightfully blamed for the result, because it followed from this failure to properly exercise our freedom. But if we are forced to give up the principle that ought-implies-can, the moral law still dictates the actions that we ought or ought not perform. But when we neglect to perform these actions, this does not represent a failure to exercise our freedom. For there was nothing we could have done to alter the order of things that was presented to us in intuition. And when we are blamed for something that ought to have happened (but did not), or something that ought not to have happened (but did), we are being blamed for something we had no power to prevent, because it did not follow from a failure to exercise our freedom. Prima
facie, this disobedience to the moral law should not be imputed to us, because our disobedience was not freely willed by us.

**The Imputation Problem**

The problem of imputation is concerned with precisely this sort of disconnect between the moral law and moral responsibility. Beck describes the problem of imputation in the following terms: “Consider the following dilemma which has embarrassed many defenders of Kant, who have accepted the conclusion that a good will is a free will. If there is evil, it must be a result of a failure to be free. Therefore, either there is no moral evil, all evil being natural and therefore not imputable to human responsibility, or goodness of will is not equivalent to moral freedom” (Beck 1960, p. 203). The reason that there is a problem of imputation is because Kant maintains, as a central precept of his practical philosophy, that a free will and a moral will are one and the same. To this end, we find him writing in the *Groundwork*: “… Since the concept of causality brings with it that of laws in accordance with which, by something that we call a cause, something else, namely an effect, must be posited, so freedom, although it is not a property of the will in accordance with immutable laws but of a special kind; for otherwise a free will would be an absurdity” (4:446). By defining the will as a causality, Kant assumes that there is a law that determines the will and the effects that follow its exercise (4:446). The causality of a will that is naturally necessitated is determined by the law of nature and thereby not free. For Kant, the only other law that can possibly determine the causality of the will is the moral law, hence he concludes that the free will is a moral will because it must be determined “in accordance with immutable laws but of a special kind” (4:446). In having its causality be determined by the moral law, it seems that the free will can only be the cause of moral actions. And this means that moral evil is not to be explained by a will that is determined by the moral law, but by a will that is determined by the
laws of nature. Moral evil, as Beck argued, would therefore be the product of natural necessity, and not freedom.

But Beck was hardly the first to point out this problem. In Anglo-American literature, Sidgwick was influential in his objection to this identification of the good will and the free will in his paper, “The Kantian Conception of Free Will” (1888). There, he accuses Kant ofequivocating between two notions of freedom. The first notion, which Sidgwick calls Good Freedom, is “the Freedom that is only realized in right conduct, when reason successfully resists the seductions of appetite or passion” (Sidgwick 1888, p. 405). The second notion, which Sidgwick calls Neutral Freedom, is “the Freedom to choose between right and wrong, which is, of course, equally realized in either choice” (Sidgwick 1888, p. 405). According to Sidgwick, when Kant identifies the will as pure practical reason, “this identification obviously excludes the possibility of Will’s choosing between Reason and non-rational impulses”, because pure practical reason, as noumena, is undetermined by empirical conditions and therefore also undetermined by sensible desires (Sidgwick 1888, p. 411). The result is that “the will, qua free, acts in accordance with these [moral] laws; – the human being, doubtless, often acts contrary to them; but then, according to this view, its choice in such actions is determined not “freely” but “mechanically”, by “physical” and “empirical” springs of action” (Sidgwick 1888, p. 411). As such, Sidgwick thinks that Kant must either give up the idea of the heteronomy of the will, in which the will yields to empirical or sensible impulses, or he must give up the idea that we can maintain moral responsibility and moral imputation for evil acts (Sidgwick 1888, p. 412). And this conclusion of Sidgwick’s sounds very

---

6 Allison records Reinhold as one of the earliest critics of Kant’s identification of the good will and the free will: “In the second volume of his Letters on the Kantian Philosophy (Briefe über die Kantische Philosophie), Reinhold, like many subsequent critics, objected to Kant’s identification of the will with practical reason and freedom in the positive sense with autonomy on the grounds that it makes it inconceivable how one could freely violate the moral law” (Allison 1990, p. 133).

7 For a two-aspect interpretation spin of the problem of imputation, consider Sullivan: (1) Kant’s transcendental idealism is best understood in terms of a two-aspect interpretation “in order to defend the claim that
much like the dilemma faced by the defenders of Kant who have “accepted the conclusion that a
good will is a free will” (Beck 1960, p. 203).

This interpretation of Kant is not without textual support. As we had mentioned earlier,
Kant is adamant in the *Groundwork* that the freedom of the will entails “a causality in accordance
with immutable laws but of a special kind” (4:446). Insofar as the laws of nature are specifically
excluded by a will that has freedom as a property, because “freedom would be that property of
such causality that it can be efficient independently of alien causes determining it” (4:446), it
follows that these immutable laws that Kant is referring to can only be the laws of reason (the
moral law). And this is taken as an indication that Kant himself accepts the conclusion that “a good
will is a free will” (Beck 1960, p. 203). But the most suggestive passages for this interpretation are
found in the Second Critique, where Kant argues that the concept of the moral law entails the
concept of freedom, and *vice versa*. Allison has named this the Reciprocity Thesis: “This is the
claim that morality and freedom are reciprocal concepts, henceforth termed the Reciprocity Thesis.
Its significance stems from the fact that it entails that freedom of the will (transcendental freedom)
is not only a necessary but also a sufficient condition of the moral law” (Allison 1990, p. 201).
The argument of the Reciprocity Thesis is divided into two stages. The first stage, under the
heading Problem I, tries to show that a moral will would also have to be a free will. This aims to

---

freedom (and so also morality) is at least *not incompatible with nature*” (Sullivan 1989, p. 281). (2) “For our theoretical
purposes we must take the viewpoint that every phenomenal agent is completely bound by causal natural laws; for our
moral purposes, we must take the viewpoint that all human agents have pure practical reason, enabling them to be
transcendently free” (Sullivan 1989, p. 281). (3) Therefore, heteronomous action is seen from the viewpoint of
phenomena, since “we must regard that agent as being causally determined as the movements of a clock” (Sullivan
1989, p. 282). And autonomous action is seen from the viewpoint of noumenal, because “an act of freedom cannot
(lie a natural effect) be derived and explained according to the natural law of the connection of effects with their
causes, all of which are appearances” (Sullivan 1989, p. 281). The idea seems to be that, since we must regard the
noumenal as the ground of the phenomenal, it is inconsistent to regard the noumenal agent as “influenced by
phenomenal desires so as to act immorally, thereby losing noumenal freedom” (Sullivan 1989, p. 283). But it seems
the following problem also presents itself: If, as the two-aspect interpretation claims, *every* action is causally-
determined from the phenomenal viewpoint (heteronomy) and free from the noumenal viewpoint (autonomy), then it
follows that *every* action would be heteronomous behavior from the phenomenal viewpoint and autonomous behavior
from the noumenal standpoint, rather than being a case of *either* heteronomy or autonomy.
prove that the moral law is a sufficient condition for freedom. The second stage, under the heading Problem II, tries to show that a free will would also have to be a moral will. This, in turn, proves freedom to be a sufficient condition for the moral law. Taken together, the results of Problem I and II show that freedom and the moral law are reciprocal concepts. Therefore, if Kant’s argument in the Second Critique succeeds as intended, then it seems to follow immediately from the Reciprocity Thesis that “a good will is a free will” (Beck 1960, p. 203).

**Spontaneity**

What does it mean for the human being to be free in the sense of spontaneity? In the First Critique, Kant describes freedom in the following way:

*Freedom in the practical sense* is the independence of the power of choice from *necessitation* by impulses of sensibility. For a power of choice is *sensible* insofar as it is *pathologically affected* (through moving-causes of sensibility); it is called an *animal* power of cause (*arbitrium brutum*) if it can be *pathologically necessitate*. The human power of choice is indeed an *arbitrium sensitivum*, yet not *brutum* but *liberum*, because sensibility does not render its action necessary, but in the human being there is a faculty of determining oneself from oneself, independently of necessitation by sensible impulses (A534/B562).

Now if it is assumed that there is no causality other than that in accordance with the laws of nature, it then follows that the human being’s power of choice is determined solely in accordance with the laws of nature. This power of choice would be pathologically necessitated, because the “moving-causes of sensibility” are necessary and sufficient to determine the power of choice. The human being’s power of choice would therefore have to be an *arbitrium brutum*. But if it is assumed that the human being’s power of choice could be determined by something other than the “moving-causes of sensibility,” then the power of choice would not be pathologically *necessitated* but merely pathologically *affected*. And a power of choice that is pathologically affected would not be an *arbitrium brutum*, but an *arbitrium liberum*. According to Kant, the human being’s power of
choice could indeed be determined by something other than the “moving causes of sensibility”, because in the human being is a faculty that is capable of “determining oneself from oneself, independently of necessitation by sensible impulses” (A534/B562). And because the human being has such a faculty, this entails that his power of choice can be determined by something other than the “moving-causes of sensibility”. The human being is therefore an arbitrium liberum, and not an arbitrium brutum.

The concept of a maxim is given in two separate locations in the *Groundwork*. It is explained first in Section I, where Kant writes: “A maxim is the subjective principle of volition; the objective principle (i.e., that which would also serve subjectively as the practical principle for all rational beings if reason had complete control over the faculty of desire) is the practical law” (4:400n). It is then explained again in Section II, where Kant writes: “A maxim is the subjective principle of acting, and must be distinguished from the objective principle, namely the practical law. The former contains the practical rule determined by reason conformably with the conditions of the subject (often his ignorance or also his inclinations), and is therefore the principle with

---

8 For an overview of the different interpretations of what maxims are, see Gressis in “Recent Work on Kantian Maxims I” and “Recent Work on Kantian Maxims II”. This paper has remained relatively ambivalent on the nature of maxims, but the relevant interpretations to consider would be Bittner and Höffe’s interpretations of maxims as Lebensregeln and Allison’s interpretation of maxims as hierarchically arranged principles. According to Bittner and Höffe, maxims are Lebensregeln (life-rules). This means that they are general principles that “denote the manner in which one lives his life as a whole” (Gressis 2010a, p. 219). And Gressis indicates that this conception of maxims as Lebensregeln gives them a particular form: “So, maxims are principles of the form, ‘when in general circumstances C, I should perform action A’ adopted on the basis of an outlook like, ‘the world is M’, where M is a descriptive evaluation (e.g., ‘a bad place for people who are nice’)” (Gressis 2010a, p. 219). According to Allison, maxims are hierarchically arranged principles. To be precise, he writes: “A maxim may be characterized as a self-imposed, practical principle or rule of action of the form: When in S-type situations, perform A-type actions” (Allison 1990, pp. 89-90). Because these S-type situations can be more or less general, this leads Allison to conclude that maxims can be “arranged hierarchically, with the more general embedded in the more specific” (Allison 1990, p. 93). These interpretations of the nature of maxims are proffered because they both consider maxims to be principles which have several rules under them. And this makes them a better fit for the Religion’s introduction of the fundamental maxim of an agent: “Now, the term “deed” can in general apply just as well to the use of freedom through which the supreme maxim (either in favor of, or against, the law) is adopted in the power of choice, as to the use by which the actions themselves (materially considered, i.e. as regards the objects of the power of choice) are performed in accordance with that maxim. The propensity to evil is a deed in the first meaning (peccatum originarium), and at the same time the formal ground of every deed contrary to law according to the second meaning…” (6:31).
which the subject acts; but the law is the objective principle valid for every rational being, and the principle in accordance with which he ought to act, i.e., an imperative” (4:421n). In both cases, the concept of a maxim is contrasted with the concept of a practical law as a subjective, as opposed to an objective, principle. But, according to Allison, there is a significant difference between the two formulations that is not readily apparent.

The formulation in 4:400n describes the practical law as an objective principle that could “also serve subjectively as the practical principle for all rational beings if reason had complete control over the faculty of desire” (4:400n). In this formulation, both maxims and practical laws are considered to be first-order principles, as Allison observes: “The first suggests that objective practical principles could also be subjective and, a fortiori, serve as maxims, at least for that subset of rational agents in whom reason is fully in control” (Allison 1990, p. 87). The formulation in 4:421n describes the maxim as the principle by which we act, in contradistinction to the practical law as the principle by which we ought to act. In this formulation, maxims are exclusively first-order principles, whereas practical laws are exclusively second-order principles. Allison provides two reasons for preferring the second formulation. First, he observes that the maxim is subjective

---

9 Working off the distinction between maxim and practical law given in 4:400n, Beck has proposed that Kant had in mind “a trichotomy, not a dichotomy, to wit: (a) mere maxim, (b) law, (c) law which is also a maxim” (Beck 1960, p. 82). Note that in the first formulation, the distinction between subjective and objective consists in whether it is a principle acted on by a subject, or a practical principle that all rational subjects can in principle act on, if reason had complete control over the faculty of desire (Allison 1990, p. 87). From this, we can establish the following possibilities, according to Beck. The first possibility is this: “A rational man acting according to some maxim while holding it to be valid only for him, and for him only because its condition is the actual state of his own motives” (Beck 1960, p. 81). This possibility represents the mere maxim. The rational man has no difficulty acting on the practical principle, but the practical principle is not one that can “also serve subjectively as the practical principle for all rational beings if reason had complete control over the faculty of desire” (4:400n). The second possibility is this: “A rational man recognizing a condition valid for, though not necessarily effective in, all rational beings” (Beck 1960, pp. 81-82). This possibility represents the practical law. The rational man recognizes the practical principle he is supposed to act on, but due to his subjective condition, finds himself unable to act on it. The last possibility is this: “A rational man recognizing a condition as present and effective in all rational beings as such and therefore as valid for and applicable to himself” (Beck 1960, p. 82). This possibility represents the practical law that is also a maxim. The rational man recognizes the practical principle he is supposed to act on, and on this basis, he adopts the practical principle as his maxim of action.
in a way that involves a condition of the subject (ignorance or inclination) in the second formulation, and this is a “distinct and stronger sense of subjectivity than the one operative in the first formulation, where it means simply acted upon by a subject” (Allison 1990, p. 87). Second, he observes that when Kant identified the moral law with an imperative, this precludes the practical law from serving as a maxim: “Imperatives, whether hypothetical or categorical, are of the wrong logical type to be maxims. They are second-order principles, which dictate the appropriate first-order principles (maxims), rather than themselves being maxims. Consequently, according to this formulation, maxims could conform (or fail to conform) to imperatives or practical laws, but they could never themselves become such laws” (Allison 1990, p. 89).

Allison develops his distinction between maxim and practical law based on the formulation given in 4:421n. In this formulation, maxims can only be first-order principles, and the practical law can only be a second-order principle. Allison’s argument hinges on Kant’s identification of the practical law with imperatives in the 4:421n formulation. For Kant defines an imperative (or command) in the following way: “All imperatives are expressed by an ought and indicate by this the relation of an objective law of reason to a will that by its subjective constitution is not necessarily determined by it (a necessitation). They say that to do or to omit something would be good, but say it to a will that does not always do something just because it is represented to it that it would be good to do that thing” (4:413). From this, it follows that imperatives are prescriptive. They say that to do or to omit something would be good. But, as Kant notes, they say it to a will that does not always do something just because it is represented to the will that it would be good to do or to omit something – they are principles that prescribe ways in which a rational subject ought to act and not the principles according to which the rational subject actually acts (4:421n). The basis for their prescriptions depends on the nature of the imperative that is involved. If the
imperative is hypothetical, then it will prescribe only those actions that “would be good merely as a means to something else” (4:414). Whereas an imperative that is categorical would prescribe actions that are “represented as in itself good” (4:414). Imperatives, therefore, provide principles for evaluating maxims (subjective principles of actions) with an eye toward certain criteria of goodness (as either being good merely as a means to something else or good in themselves). Ipso facto, imperatives are principles of the second-order.

Allison’s reason for calling the practical law a second-order principle is applicable to both imperatives that are hypothetical and imperatives that are categorical. But a practical law is an objective principle, and as “the objective principle valid for every rational being”, it can only be a categorical imperative. Kant provides the derivation of the categorical imperative in the following passage from the Groundwork:

When I think of a hypothetical imperative in general I do not know beforehand what it will contain; I do not know this until I am given the condition. But when I think of a categorical imperative I know at once what it contains. For, since the imperative contains, beyond the law, only the necessity that the maxim be in conformity with this law, while the law contains no condition to which it would be limited, nothing is left with which the maxim of action is to conform but the universality of a law as such; and this conformity alone is what the imperative properly represents as necessary.

The derivation proceeds from the concept of the categorical imperative (and compares it to the concept of a hypothetical imperative). The hypothetical imperative represents that an “action would be good merely as a means to something else” (4:414). If this something else (the end)
was not considered to be good, then the means for achieving this end would not be considered
good either. The end is therefore the condition for the means to be considered good, and since this
end is completely contingent, we cannot know what a hypothetical imperative will prescribe
without first knowing what is the end that has been adopted. Furthermore, because this end is
completely contingent, not all rational subjects will have the same ends, and anything prescribed
by a hypothetical imperative to a rational subject would not be binding unless he has adopted the
relevant end.

The categorical imperative is a different story. In a categorical imperative, “the action is
represented as in itself good” (4:414). What Kant means by this is that the action is not conditioned
on anything else as a means to an end – it is unconditionally good. And this, he reasons, entails
that “nothing is left with which the maxim of action is to conform but the universality of the law
as such” (4:421). But what is the basis for Kant’s thinking that the unconditional nature of a
categorical imperative is captured by the demand that the maxim conform to the universality of a
law as such? Kant seems here to subscribe to the idea that, since the hypothetical imperative is
conditionally practically binding on the basis of empirically given ends, then abstracting from any
and all empirically given ends would leave us with an imperative that is unconditionally practically
binding. This is because, having abstracted from any and all empirically given ends, all that

1. Whoever wills an end also wills the sole means which are in his power if he is fully rational.
2. Whoever wills an end ought to will the sole means which are within his power. Or, in other words, if a person
   wills an end and certain means are necessary to achieve that end and are within his power, then he ought to
   will those means.
3. A is the sole means to B and is generally available (within everyone’s power).

Therefore:
4. If one wills B, one ought to will A.
5. Q (a person) wills B.

Therefore:
6. Q ought to will A” (Hill 1973, pp. 431-432).
remains is a principle that is practically binding, that is, has the character of being universal and necessary. The categorical imperative bids maxims to conform with the law by commanding that we act only on those maxims that have the same character of being universal and necessary, hence Kant writes: “There is, therefore only a single categorical imperative and it is this: act only in accordance with that maxim through which you can at the same time will that it become a universal

---

11 Aune and Wood both argue that a gap arises at this point in Kant’s derivation of the categorical imperative. They say that Kant illegitimately moves from the requirement that the maxim has to conform to “the universality of a law as such” to the requirement to “act only on that maxim through which you can at the same time will that it should become a universal law”, that is, the requirement that the maxim conform to the categorical imperative. As Allison explains it, “In order for the gap to be filled, we would need a premise to the effect that a maxim that conforms to universal law just in case that maxim can include itself as universal law. But whereas a maxim that can include itself as universal law, for example, the maxim of honesty, clearly conforms to universal law, the converse does not hold. A maxim might very well conform to universal law (indeed, a law produced through the generalization of a maxim) without being able to include itself as universal law” (Allison 1991, p. 8). The example of a maxim that conforms to universal law without being able to include itself as universal law that Allison had in mind is the example of false promising (Allison 1991, p. 8). The maxim of false promising has “universality of applicability” insofar as it is perfectly rational for everyone to act on that principle and make a false promise in order to escape a difficult situation. But the maxim of false promising cannot include itself as universal law – because the advantage that the false promise was intended to secure would never materialize if the principle was universally adopted. Therefore, a gap between the weaker requirement of universality of applicability and the stronger requirement that maxims be able to include themselves when regarded as universal laws arises in Kant’s derivation of the categorical imperative.

In response, Mariña has pointed out that meta-level principles like Kant’s categorical imperative and Wood’s requirement that maxims possess universality of applicability are both capable of selecting between maxims, but only the former can correctly be called unconditioned, whereas the latter only seems unconditioned. The categorical imperative dictates that we are to act only in accordance with those maxims that we can at the same time will that it become universal laws (4:421). It is concerned with whether a contradiction would arise, should everyone adopt a given maxim as their principle of action at the same time. In the case of false promising, where a person in financial difficulty makes a false promise in order to secure a loan, this does not happen, because “the universality of a law that everyone, when he believes himself to be in need, could promise whatever he pleases with the intention of not keeping it would make the promise and the end one might have in itself impossible, since no one would believe what was promised but would laugh at all such expressions as vain pretenses” (4:422). This may be contrasted with the requirement that maxims possess universality of applicability. According to this requirement, we are to act only in accordance with those maxims that are rational for any other agent to adopt in relevant circumstances. But the universality requirement of the categorical imperative is unconditioned in a way that the universality requirement of applicability is not: “Yet while the stronger universalizability requirement is capable of throwing out those maxims having only conditioned worth, the weaker is not, but only throws out maxims that do not specify means adequate to their end. This tells us that the former requirement is independent of given ends adopted on account of sensuously conditioned desires in a way that the latter is not. In fact, the latter requirement cannot select among ends precisely because it is not really unconditioned but only seems so; the given end upon which it depends corresponds to a second-order desire, that is, the desire to fulfill one’s desires. It thus has the form of a hypothetical imperative: if you desire to fulfill your desires, then your means should be objectively practically sufficient to your end. The fact that it depends on such a second order desire is what makes it seem unconditioned, since adoption of the principle would not seem to depend upon any particular end” (Mariña 1998, pp. 176-177). Thus, only the stronger universalizability requirement gives us the unconditionality necessary for a categorical imperative, whereas the weaker universalizability requirement remains conditional in nature, and therefore has only the form of a hypothetical imperative.
And this makes the categorical imperative a principle that is valid for every rational being, i.e. objective.

The human being is an arbitrium liberum if he is pathologically affected to action by his incentives, but not pathologically necessitated to action by them. Kant expands on his account of how the power of choice is determined to action through a maxim with the introduction of the Incorporation Thesis, which Kant describes in the following passage: “… Freedom of the power of choice has the characteristic, entirely peculiar to it, that it cannot be determined to action through any incentive except insofar as the human being has incorporated it into his maxim (has made it into a universal rule for himself, according to which he wills to conduct himself)” (6:24).

According to the Incorporation Thesis, the power of choice is determined to action through the adoption of a maxim. The mere presence of an incentive does not determine the power of choice to action, but must be incorporated into the maxim in order for it to be efficacious. As a matter of fact, the human being does not incorporate only the moral incentive into his maxim, but he incorporates the nonmoral incentive as well. Kant writes in this regard: “The law rather imposes itself on him irresistibly, because of his moral predisposition; and if no other incentive were at work against it, he would also incorporate it into his supreme maxim as sufficient determination

---

12 Baron, in her article “Freedom, Frailty, and Impurity”, has pointed out a difficulty with the Incorporation Thesis. The Incorporation Thesis, remember, states that the human being’s power of choice cannot be determined to action through any incentive unless the human being has incorporated the incentive into his maxim. This assumes that it is possible for the human being who is faced with a moral and a nonmoral incentive to choose to incorporate one (and not the other) into his maxim. Baron is skeptical that this is actually possible: “… Of course I can, on Kant’s view, decide not to do x, despite my intense desire to do x. But suppose I want to do x, but for a shameful reason. I realize that the reason is shameful, and also see that there are better reasons for doing x. Can I choose to do x, but not for the shameful reason? Think of it this way. It is part of our freedom that we are not simply buffeted about by our inclinations. But there are two ways we might do this. One is to refrain from acting as the inclination directs; this is something that we clearly can, on Kant’s view, do. The other is to act as the inclination directs, but not from that inclination – for we do not incorporate that inclination into our maxim” (Baron 2008, p. 432). Baron’s point is not that the motives from which we act are often inscrutable, and we don’t know whether it is the nonmoral or the moral incentive that we have incorporated into the maxim in a situation like this. It is that the only way for us to not act as the inclination directs is by refraining from the action completely (Baron 2008, p. 432). It does not seem possible for us to simply decide to act as the inclination directs, but not from the inclination but another one (Baron 2008, p. 432).
of his power of choice, i.e. he would be morally good. He is, however, also dependent on the incentives of his sensuous nature because of his equally innocent natural predisposition, and he incorporates them too into his maxim (according to the subjective principle of self-love)” (6:36).

It follows from this that every maxim is identical with respect to their material (6:36). And because the moral incentive and the nonmoral incentive have both been incorporated into every maxim, they cannot form the basis for distinguishing between the maxims that are good and the maxims that are evil.\(^\text{13}\)

The moral bent of the maxim, whether it is good or evil, is determined by the way the moral incentive and the nonmoral incentive are related to each other, that is, by the form of the maxim and not the material of the maxim. Kant therefore writes: “Hence the difference, whether the human being is good or evil, must not lie in the difference between the incentives that he incorporates into his maxim (not in the material of the maxim) but in their subordination (in the form of the maxim): which of the two he makes the condition of the other” (6:36). If it is the nonmoral incentive that is subordinated to the moral incentive, then the maxim is considered morally good. And if it is the moral incentive that is subordinated to the nonmoral incentive, then the maxim is considered morally evil. Because the human being naturally incorporates both the moral incentive and the nonmoral incentive, it follows that one of the incentives would necessarily

---

\(^{13}\) Note that the discussion of the Incorporation Thesis in 6:24 makes no mention of the human being naturally incorporating both the moral incentive and the nonmoral incentive. There, the agreement or disagreement of the power of choice with the moral law seems to depend on whether the moral incentive is incorporated into the maxim. Insofar as the moral law is itself an incentive, disagreement of the power of choice with the moral law requires “a positive ground antagonistic to the good, = -a” (6:22n). Hence Kant writes: “Now, 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)” (6:24). From this, it seems that either the moral incentive or the nonmoral incentive is incorporated into the maxim in a spontaneous act of reason, and never both. And this is because Kant’s ethical rigorism dictates that a maxim is either good or it is bad, and never partly good and partly bad (6:24).
have to be subordinated to the other in the process of incorporation, otherwise the maxim would have no moral bent: “... Since he naturally incorporates both into the same maxim, whereas he would find each, taken alone, of itself sufficient to determine the will, so, if the difference between maxims depended simply on the difference between incentives (the material of the maxims), namely, on whether the law or the sense impulse provides the incentive, he would be morally good and evil at the same time – and this is a contradiction” (6:36). That is, if the mere act of incorporating the moral incentive into the maxim makes the maxim morally good, and the mere act of incorporating the nonmoral incentive into the maxim makes the maxim morally evil, then the fact that the human being naturally incorporates both the moral incentive and the nonmoral incentive into the maxim would make the maxim morally good and morally evil at the same time. And this, Kant argues, is a contradiction.

The point of the Incorporation Thesis is to restrict the effect incentives can have in determining the power of choice to action. According to Allison, it does so in the following way:

First, it makes it clear that for Kant an inclination or desire does not of itself constitute a reason for acting. It can become one only with reference to a rule or principle of action, which dictates that we ought to pursue the satisfaction of that inclination or desire. Moreover ... the adoption of such a rule cannot itself be regarded as the causal consequence of the desire or, more properly, of being in a state of desire. On the contrary, it must be conceived as an act of spontaneity on the part of the agent” (Allison 1990, p. 40).

For Kant, the incentive can be the cause of neither the maxim nor the action. It is not the cause of the action (or does not determine the power of choice to action), because it is the maxim that

14 Note that Frierson disagrees with Allison on this point. Frierson argues that the incorporation thesis holds only for the higher desires. With respect to the lower desires, however, he writes: “For Kant, contra Allison, human beings can, sometimes, act purely from instinct or inclination, without incorporating such instincts or inclinations into any principles of the understanding. Kant’s language to describe such “actions” fits the lack of true agency implied by their failure to fit Allison’s account of incorporation. He refers to them as actions proceeding from “stimuli” or “impulse”. Most actions, even those that are not guided by morality, are free in the sense that they are associated with the higher faculty of desire, where one acts on principles or maxims, even if these maxims take the satisfaction of inclination as their end. But one can also “act” directly from lower desires” (Frierson 2005, p. 16n).
determines the power of choice to action. In order to be efficacious, the incentive first has to be made into a rule or principle of action, that is, a maxim. And the incentive is not the cause of the maxim either, because it has to be made into a rule or principle of action, and this construction is a spontaneous act on the part of the rational agent. The rational agent incorporates both the moral incentive and the nonmoral incentive into his maxim. During this process of constituting the material of the maxim, it becomes necessary for him to determine its form: Either the moral incentive or the nonmoral incentive must be subordinated to the other. The rational agent has to either incorporate the nonmoral incentive as his condition for complying with the moral law (subordinating the moral incentive) or incorporate the moral incentive as his condition for complying with the law of self-love (subordinating the nonmoral incentive) (6:36). These acts are spontaneous, and represent the absolute beginning of a state and the series of its consequences by the Willkür. And because a maxim and the action that proceeds from it are morally good or bad depending on the incentive that is incorporated (or subordinated) in the construction of the maxim, the spontaneity that the rational agent exercises to decide which incentive to incorporate (or subordi-nate) in the maxim is freedom in the sense relevant to the assignment of moral responsibility and moral imputation.

**Autonomy and Heteronomy**

In the *Groundwork*, Kant describes the will as having the property of being heteronomous or the property of being autonomous. The will has the property of being heteronomous in the following circumstances: “If the will seeks the law that is to determine it anywhere else than in the fitness of its maxims for its own giving of universal law – consequently if, in going beyond itself, it seeks this law in a property of any of its objects – heteronomy always results” (4:441). This indicates that the will has the capacity to give itself the law, but it fails to exercise this capacity
when it places itself in subservience to objects that determine it according to the laws of nature. In so doing, the will gives up its independence from nature, and this is what makes the heteronomous will unfree. As Beck puts it, “The will can “obey” laws of nature in the sense of intentionally using them in the form of means-end statements, in seeking men’s goals in nature” (Beck 1960, p. 196). This is contrasted with the autonomy of the will, which Kant describes so: “Autonomy of the will is the property of the will which is a law to itself (independently of any of the objects of volition)” (4:440). When the will exercises its capacity to give itself the law, this is an exercise of the Wille, which Kant uses to refer to the legislative aspect of the will in his later writings. This is described by Beck in the following passage: “It gives a law to Willkür which it has freely legislated – legislated by the necessity of its own nature. The faculty of lawgiving is will in the sense in which pure practical reason is will, and its legislation is “freedom in the positive sense”, or autonomy” (Beck 1960, p. 197). This act of self-legislation can be contrasted with heteronomy, where the will allows itself to be legislated to by outside sources. In being “legislated by the necessity of its own nature”, the will accepts no outside source for its laws, and this self-determination makes the will autonomous, that is, free.

For an example of heteronomy, let us turn to Hume’s theory of moral motivation. Hume, who had famously declared that reason is, and ought to be, only the slave of the passions, can be considered a representative of the passions-based theory of moral motivation (Treatise 2.3.3.4). For Hume, the power of choice can only be determined by our desires and inclinations, which he

---

15 Note that there is disagreement in the literature on this point. Allison does not consider the freedom of the Wille to be autonomy, as we can see in the following passage: “... It is Wille in the narrow sense that provides the norm and Willkür that chooses in light of this norm. Similarly, this distinction allows us to speak of the will as giving the law to, or even as being the law for, itself, since this is just a matter of Wille giving the law to, or being the law for, Willkür. Strictly speaking, then, it is only Wille in the broad sense that can be characterized as a law to itself” (Allison 1990, pp. 130-131). When Beck argues that Wille, as the legislative aspect of the will, has freedom in the sense of autonomy, he has in mind what Allison would call “Wille in the narrow sense”. Allison, on the other hand, uses “Wille in the broad sense” to denote the will, and this is subdivided into an executive aspect (Willkür) and a legislative aspect (Wille in the narrow sense).
labels the passions in the following passage: “Nothing can oppose or retard the impulse of passion, but a contrary impulse; and if this contrary impulse ever arises from reason, that latter faculty must have an original influence on the will, and must be able to cause, as well as hinder any act of volition” (Treatise 2.3.3.4). Reason has no influence over our power of choice because Hume considers the province of reason to be restricted to demonstration and probability (Treatise 2.3.3.2). But demonstration has to do with the relation of ideas, and therefore “its proper province is the world of ideas, and as the will places us in that of realities, demonstration and volition seem, upon that account, to be totally remov’d, from each other” (Treatise 2.3.3.2). In other words, the abstract reasoning involved in mechanics and mathematics have no place in determining the power of choice. Likewise, for Hume, probability has to do with the relation of cause and effect. Therefore, probabilistic reasoning does nothing to determine our power of choice either. This is because merely knowing that an object is related to another as cause to effect is not sufficient to bring about action. It is only when these causal connections are further connected to the feelings of pleasure and pain that we can be determined to action as a result. But absent this connection to the feelings of pleasure and pain, we would be completely indifferent to the knowledge that a particular object is related to another as cause to effect (Treatise 2.3.3.3).

Because Hume thinks the power of choice can only be determined by our passions, and not by reason, it also follows that reason is powerless to prevent the determination of the will by the passions. He argues this point by means of a simple denial of the consequent: “Nothing can oppose or retard the impulse of passion, but a contrary impulse; and if this contrary impulse ever arises from reason, that latter must have an original influence on the will, and must be able to cause, as well as hinder any act of volition. But if reason has no original influence, ’tis impossible it can withstand any principle, which has such an efficacy, or even keep the mind in suspense a moment”
(Treatise 2.3.3.4). Insofar as the will can only be determined by an original influence, it follows that passion has an original influence on the will that allows it to determine the will. Likewise, it follows that reason would need to be an original influence in order to similarly determine the will. But it is not. From Hume’s earlier arguments, reason is restricted to demonstrating the relation between ideas and relating objects to each other as cause to effect, neither of which provides reason with any original influence on the will. Therefore, reason does not have any way to prevent the determination of the will by the passions. The only way to prevent the determination of the will by a passion is by counteracting its original influence with the original influence of another passion. If the power of choice has been determined in a particular direction, the only way to determine the power of choice otherwise is by means of a contrary passion that determines the power of choice in the opposite direction. In a situation like this, where the power of choice is being pulled in different directions by opposing passions, the direction in which the power of choice ends up heading is a function of the passion with the stronger pull. It is for this reason that the metaphor most often used to describe a passions-based theory of moral motivation like Hume’s tends to appeal to the sum of vector forces or of mechanical forces acting on the object.

There are strong parallels between this and Kant’s account of the lower faculty of desire. According to Kant, the faculty of desire rests upon the following principle: I desire nothing but what pleases, and avoid nothing but what displeases (29:894).\(^\text{16}\) The desire for an object is always connected to the pleasure we take in the object, and aversion for an object is always connected to

\(^{16}\) For a comprehensive account of the empirical side of Kant’s moral psychology, see Frierson’s “Kant’s Empirical Account of Human Action” (2005). In particular, he notes that Kant’s concept of desire is not identical to the concept of desire as it is used in Anglo-American philosophy: “Because desires simply are representations insofar as those representations are directed towards action, Kant’s notion of desire is more closely connected to choice and action than the customary English sense of desire, whereby one can desire something without actually pursuing it. Once one has a desire in this general sense, one is committed to action, and action follows necessarily in the absence of unforeseen hindrances. One might, for example, desire a mango and then find oneself unable to climb the tree, but one’s representation will not count as a desire unless it prompts one to action. In contrast to customary English usage, Kantian desires mark an end to deliberation, not factors taken into account in deliberation” (Frierson 2005, p. 10).
the displeasure we take in the object. The desires themselves are divided by reference to their origins in cognition. The ones Kant calls the lower desires have their source in sensibility, whereas the ones Kant calls the higher desires have their source in the understanding, either in whole or in part. The passage where Kant provides a detailed explanation of these distinctions is in the Metaphysik $L_2$:

The intellectual impelling cause $<causa impulsiva intellectualis>$ is either purely intellectual without qualification $<simpliciter talis, mere intellectualis>$, or in some respect $<secundum quid>$. When the impelling cause $<causa impulsiva>$ is represented by the pure understanding, it is purely intellectual $<mere intellectualis>$, but if it rests on sensibility, and if merely the means for arriving at the end are represented by the understanding, then it is said to be in some respect $<secundum quid>$ (28:589).

From this passage, we can ascertain that there are three types of desires for Kant. Lower desires have their source entirely in sensibility, and the understanding plays no part in the formation of these desires. The lower desires are moved by the stimuli of the senses, and a power of choice that is moved in this way is the brute power of choice ($arbitrium brutum$). The higher desires can be divided into those that are purely intellectual and those that are mixed, that is, desires that are partly intellectual and partly sensible. Higher desires that are purely intellectual have their source entirely in the understanding, and sensibility plays no part in the formation of these higher desires. As such, the ends that are set by a higher desire that is purely intellectual have to be objective ends. But higher desires that are merely “intellectual in some respect” have their source in both sensibility and the understanding. The ends that are desired are provided by sensibility, like in the

\[\text{Note that the converse is not true. According to Kant, it is possible to have a feeling without having the corresponding desire: “... There can be a pleasure that is not connected with any desire for an object but is already connected with a mere representation that one forms of an object (regardless of whether the object of the representation exists or not)” (6:211). The exception that Kant is making here has to do with the case of aesthetic pleasure. The pleasure we take in a beautiful sculpture is unrelated to our desire for the sculpture, that is, our desire to possess the sculpture. And it would indeed be possible to take pleasure in the beauty of the same sculpture if it only existed in our imagination (5:204-205).}\]
lower desires. And because the ends that are desired are produced by sensibility, these ends are subjective ends, and hold only for the subject. But understanding plays a role in determining the means to achieving these ends. The fact that the understanding plays a role in the formation of these desires leads Kant to classify them as higher desires, rather than lower desires. And this makes Kant’s account of higher desires that are partly intellectual and partly sensible reminiscent of Hume’s declaration that reason is the slave to the passions: “Reason is, and ought only to be the salve of the passions, and can never pretend to any other office than to serve and obey them” (Treatise 2.3.3.4).

Reason is subservient to the passions when the ends of an action are determined by passion, and reason prescribes the means by which these ends may be achieved. Because the ends of an action are determined by passion, it may be characterized in terms of receptivity: “Pleasure arising from the representation of the existence of a thing, insofar as it is to be a determining ground of desire for this thing, is based on the receptivity of the subject, since it depends upon the existence of an object…” (5:22). Insofar as the passions depend on the receptivity of the subject, there must be an external object to inflame the passions. Absent an object to be passionate about, there would be no occasion for reason to prescribe anything to the subject. In addition, the presence of an

18 Mariña notes that a shift occurs in Kant’s attitude towards these material practical principles. These are principles that are “constructed in order to achieve given desired ends; as such their content depends upon those ends. Their adoption thus hinges upon the existence of contingent desires making up what Kant dubs “the lower faculty of desires” (Mariña 2001, p. 184). Now Kant had, in the *Groundwork*, declared actions having their origins in the inclinations to be without moral worth (4:398). By the Second Critique, Kant had moved beyond this position, stating: “All material practical principles as such are, without exception, of one and the same kind and come under the general principle of self-love or one’s own happiness” (5:22). This means that the actions of the sympathetically-attuned soul from the *Groundwork* are not only without moral worth, but also done out of self-love or his own happiness. Mariña explains the basis for making this stronger claim in the following way: “... The issue rather lies in that in acting in such a way one makes one’s causal history the basis of the ultimate criterion determining one’s choices. In doing so one makes the self and its specific particulars the locus of the determination of what is of value. To act on the principle of satisfying those desires one simply happens to have is to act on the principle of self-love because to do so is to validate the choices of the self qua particular individual. Attitudes toward the other stemming from these desires alone, however other-directed they may seem (as in the case of the friend of man), cannot genuinely be directed toward the other qua other, that is, as an individual having a claim on our regard toward him or her irrespective of our likes and dislikes” (Mariña 2001, p. 188).
external object would inflame no passions if the subject itself were not receptive towards it. Absent the proper internal dispositions in the subject, the subject would be incapable of being passionate about the object, and there would again be no occasion for reason to prescribe anything to the subject. These factors combined entail that the principles prescribed by a reason subservient to the passions would have to be empirical principles: “Empirical principles are not at all fit to be the ground of moral laws. For, the universality with which these are to hold for all rational beings without distinction – the unconditional practical necessity which is thereby imposed upon them – comes to nothing if their ground is taken from the special constitution of human nature or the contingent circumstances in which it is placed” (4:442). This means that, for Hume, only hypothetical imperatives are possible.

For Hume, reason is always subservient to the passions. These passions can be divided into two kinds. There are the calm passions, of which Hume writes: “Now ‘tis certain, there are certain calm desires and tendencies, which, tho’ they be real passions, produce little emotion in the mind, and are more known by their effects than by the immediate feeling or sensation” (Treatise 2.3.4.8). And these calm passions are pitted against the violent passions: “Besides these calm passions, which often determine the will, there are certain violent emotions of the same kind, which have likewise a great influence on the faculty. When I receive any injury from another, I often feel a violent passion of resentment, which makes me desire his evil and punishment, independent of all considerations of pleasure and advantage to myself” (Treatise 2.3.4.9). Both these passions can be said to seek their law in a property of their object. This is because what moves the Humean will is the prospect of pain and pleasure, and “when we have the prospect of pain or pleasure from any object, we feel a consequent emotion of aversion or propensity, and are carry’d to avoid or embrace what will give us this uneasiness or satisfaction” (Treatise 2.3.3.3). But consider the Humean will
in relation to the definition of heteronomy that Kant gives in the *Groundwork*: “If the will seeks the law that is to determine it anywhere else than in the fitness of its maxims for its own giving of universal law – consequently if, in going beyond itself, it seeks this law in a property of any of its objects – heteronomy always results” (4:441). Insofar as passion is the only influence capable of determining the Humean will in any way whatsoever, it follows that heteronomy is the only possible basis of the Humean will.

Because there is no such thing as an autonomous will for Hume, he does not consider freedom in terms of autonomy, but in terms of spontaneity and indifference: “Few are capable of distinguishing betwixt the liberty of spontaneity, as it is call’d in the schools, and the liberty of indifference, betwixt that which is oppos’d to violence, and that which means a negation of necessity and causes” (Treatise 2.3.2.1). For Hume, freedom means either liberty of spontaneity or liberty of indifference. Liberty of spontaneity has to do with the power we have to act or not act according to the determinations of our will. Therefore, an act is spontaneous (and therefore free) if it originates from internal causes rather than from an external compulsion. For Hume, the freedom that is necessary for the purposes of moral responsibility and moral imputation is liberty of spontaneity, as he makes clear in the following passage: “Actions are by their very nature temporary and perishing; and where they proceed not from some cause in the characters and disposition of the person, who perform’d them, they infix not themselves upon him, and can neither redound to his honour, if good, nor infamy, if evil” (Treatise 2.3.2.6). By contrast, liberty of indifference is defined as the power to act independently of the laws of nature – that our actions are not necessitated or caused by the laws of nature. But besides the laws of nature, Hume thinks there is no other law that is capable of determining our actions. Hence he denies liberty of indifference in the following passage: “According to my definitions, necessity makes an essential
part of causation; and consequently liberty, by removing necessity, removes also causes, and is the very same thing with chance” (Treatise 2.3.1.18).

The resemblance of this line of reasoning to Kant’s in the *Groundwork* is worth noting. Remember that Kant writes in the *Groundwork*: “Since the concept of causality brings with it that of laws in accordance with which, by something that we call a cause, something else, namely an effect, must be posited, so freedom, although it is not a property of the will in accordance with natural laws, is not in accordance with immutable laws but of a special kind; for otherwise a free will would be an absurdity” (4:446). Hume and Kant share a similar starting-point on this matter. They both agree that there is a conception of freedom that can be understood as the power to act independently of the law of nature. In addition, their shared conception of the will as a causality entails that there must be a law that determines the exercise of its causal power. But the two philosophers part ways past this point. For Hume, there is no law capable of determining the will other than the law of nature. Therefore, a will that has the power to act independently of the law of nature would be lawless, and to call such a will a *free* will would be an absurdity. Kant, by contrast, believes that a will that is not determined by the law of nature would still be determined by “immutable laws but of a special kind”, that is, the moral law. Therefore, a will that has the power to act independently of the law of nature would not be lawless – it would be a will that acts in accordance with the moral law. Such a will would possess freedom in the sense of indifference, which would be a characterization of its freedom in only the negative sense. For a will whose causality is determined by the moral law would also be free in the *positive* sense, in the sense which Kant calls *autonomy*.19

---

19 Neither Hume’s liberty of spontaneity nor his liberty of indifference can be considered autonomy. Hence, the Humean will, for whom it is impossible to prescribe pure principles in the form of the categorical imperative, is incapable of self-determination in the full sense needed for autonomy. For Hume, self-determination is compatible with being causally determined to act according to laws that govern our inner representations, and this is precisely the
The Reciprocity Thesis

We have seen earlier that there are interpreters that take the Reciprocity Thesis as establishing for Kant the conclusion that a free will is a moral will. They argue that a free will has to at least have freedom in the negative sense. 20 It is a will that is free from the influence of alien causes. The Reciprocity Thesis then uses this negative freedom to prove the independence of the will from the law of nature, which leaves only the moral law to determine it. But this negative freedom is also proof that the will has the power to do otherwise, and this being the case, it is a will that can be assigned moral responsibility. Sidgwick points out the bearing negative freedom has on moral responsibility when he discusses the “metaphysical doctrine of a double kind of causation in human actions” in the following passage: “According to Kant, every such action, regarded as a phenomenon determined in time, must be thought as a necessary result of determining causes in antecedent time – otherwise its existence would be inconceivable – but it may also be regarded in relation to the agent considered as a thing-in-itself, as the “noumenon” of which the agent so considered in relation to his phenomena” (Sidgwick 1888, p. 409). The import of this doctrine is that the law of nature is applicable only to the phenomenal agent, but the agent,
considered as a thing-in-itself, is independent from these laws. That is, the agent, considered as a thing-in-itself, is free from the influence of alien causes. This is freedom in the negative sense. It turns out, therefore, that negative freedom is at once necessary for both the identification of the free will with the moral will and also for the assignment of moral responsibility and moral imputation to the agent. And because a will to which we can assign moral responsibility and moral freedom is free whether or not it disobeys the moral law, a free will such as this cannot be identical with the moral will. Thus, it seems that Beck is right: either there is no moral evil, all evil being natural and therefore not imputable to human responsibility, or goodness of will is not equivalent to moral freedom (Beck 1960, p. 203).

But there is a way out from this unhappy conclusion, and that is by showing that the Reciprocity Thesis has to do with freedom as autonomy, that is, it asserts that autonomy and morality are reciprocal concepts. In examining Problem I, we see that Kant intends to show that the moral will has to be a free will: Supposing that the mere lawgiving form of maxims is the only sufficient determining ground of a will – to find the constitution of a will that is determinable by it alone (5:28). From our discussion of the categorical imperative earlier, we know that “the mere lawgiving form of the maxims” refers to the characteristics of universality and necessity that make a maxim lawlike when they are present in a maxim. And the categorical imperative bids us to act only on those maxims that have these characteristics, and to refrain from acting on maxims in which these characteristics are absent. Now a will that obeys the categorical imperative is a will in which the mere lawgiving form of maxims is a sufficient determining ground of a will, but this is not what Kant is interested in here – he wants to discover the constitution of a will for whom the mere lawgiving form of maxims is its only sufficient determining ground. As Hudson observes, both the matter and the form of the maxim can serve as the sufficient determining ground of the
will. But this would be a will that obeys both hypothetical imperatives (when the matter of the maxim is the sufficient determining ground of the will) and categorical imperatives (when the form of the maxim is the sufficient determining ground of the will). If the mere lawgiving form of the maxim is the only sufficient determining ground of the will, then this excludes the matter of the maxim as a possible determining ground of the will, and this would be a will that only obeys the categorical imperative (Hudson 1991, pp. 192-193).

Having shown what it means for the mere lawgiving form of maxims to be the only sufficient determining ground of a will, Kant goes on to argue that such a will has to have a particular constitution, namely that of a free will: “But if no determining ground of the will other than the universal lawgiving form can serve as a law for it, such a will must be thought as altogether independent of the natural law of appearances in their relations to one another, namely the law of causality” (5:29). By definition, a will that can be determined by the law of causality would not be a will for whom the only sufficient determining ground is the moral law. The contrast that Kant sets up is between a will that is determined by the law of causality and a will that is determined by the moral law. For a will that is determined by the law of causality has goals that are given to him by nature, because it seeks the law that is to determine it in its objects, rather than in the fitness of its maxims for its own giving of universal law. And that means a will that is determined by the law of causality is heteronomous. But a will that is determined by the moral law obeys a law that is...

---

21 Hudson puts it slightly differently: “Wille as [negatively free] can give an empirical practical law which can determine Willkür, but it is not sufficient to do so by itself; rather an antecedent desire is required as an additional necessary condition for this determination (Hudson 1991, p. 193). For Hudson, a Wille that is free is the lawgiver of both empirical practical laws and pure practical laws. In this, he follows Meerbote in subdividing the Wille into empirical practical reason and pure practical reason (Hudson 1991, p. 185 – see also Meerbote 1983, pp. 69-84). Therefore, he understands the freedom of Wille in the negative sense as “the freedom of the Wille to give empirical practical laws – also called heteronomy of the will” (Hudson 1991, p. 191). And freedom of the Wille in the positive sense is understood as “freedom of the Wille to give the moral law – also called autonomy of the will” (Hudson 1991, p. 191). Therefore, he reads Kant in Problem I as stating: If we assume freedom of the Wille in the positive sense (autonomy) and not freedom of the Wille in the negative sense (heteronomy), then the will also has freedom of the Willkür in the negative sense, that is, freedom in the transcendental sense (Hudson 1991, p. 194).
legislated by the necessity of its own (rational) nature. It seeks the law within itself, and not in an outside object. Therefore, a will that is determined by the moral law is autonomous. Problem I of the Reciprocity Thesis must therefore be interpreted as: A will for which the mere lawgiving form of a maxim can alone serve as a law is an autonomous will.

Turning to Problem II, Kant intends to demonstrate here that a free will has to be a moral will: Supposing that a will is free: to find the law that alone is competent to determine it necessarily (5:29). Freedom is defined in terms of an independence from empirical conditions in the following passage: “Since the matter of a practical law, that is, an object of maxim, can never be given otherwise than empirically whereas a free will, as independent of empirical conditions (i.e., conditions belonging to the sensible world), must never be determinable, a free will must find a determining ground in the law but independently of the matter of the law” (5:29). This definition of freedom is negative, because the assignment of moral responsibility and moral imputation is also made on the basis of the will’s independence from empirical conditions. But this would mean that negative freedom of the will is (1) the sufficient condition for assigning moral responsibility and moral imputation to its actions, and (2) the sufficient condition for being a moral will (i.e. a will that finds its determining ground in the mere lawgiving form of the law). The dilemma that Beck describes arises at this point, once we notice that a will that is free for the purposes of moral imputation need not be a moral will (i.e. it only needs to be a will that can either act morally or act immorally). But a will that finds its determining ground in the mere lawgiving form of the law has to be a moral will (i.e. it would be a will that cannot act immorally). The way out of this dilemma, however, is to realize that negative freedom is applicable to both the freedom of the Wille and the freedom of the Willkür. The Wille needs to be free in the negative sense in order to “find a determining ground in the law but independent of the matter of the law” (5:29). And the Willkür
needs to be free in the negative sense in order for the assignment of moral responsibility and moral imputation to take place.

And why does the Wille need to be free in the negative sense in order to “find a determining ground in the law but independently of the matter of the law”? Kant’s argument in Problem II relies on his earlier distinction between the matter and the form of the law. He defined the matter of the law in the following terms: “The matter of a practical principle is the object of the will. This is either the determining ground of the will or it is not. If it is the determining ground of the will, then the rule of the will is subject to an empirical condition (to the relation of the determining representation to the feeling of pleasure or displeasure), and so is not a practical law” (5:27). But if the will is a will that is independent of empirical conditions, then the rule of the will is not subject to an empirical condition (not subject to the relation of the determining representation to the feeling of pleasure and displeasure), and so the determining ground of the will is not the matter of the practical principle (or law). But besides the matter of the practical principle, there is only the form of the practical principle: “Now, all that remains of a law if one separates from it everything material, that is, every object of the will (as its determining ground), is the mere form of giving universal law” (5:27). It is the legislative aspect of the will (the Wille) that is responsible for the giving of the moral law. For this, freedom of the negative sense is necessary but not sufficient, because the act of lawgiving by the Wille is an exercise of self-determination, and this is freedom in the positive sense (or autonomy). Hence: “… The sole principle of morality consists in independence from all matter of the law (namely, from a desired object) and at the same time in the determination of choice through the mere form of giving universal law that a maxim must be capable of. That independence, however, is freedom in the negative sense, whereas this lawgiving of its own on the part of pure and, as such, practical reason is freedom in the positive sense” (5:33).
Autonomy is this freedom in the positive sense that entails a will determined by the moral law. Problem II of the Reciprocity Thesis must therefore be interpreted as: Supposing that the will is autonomous, the lawgiving form, insofar as this is contained in the maxim, is therefore the only thing that can constitute a determining ground for such a will.

Thus, the solution to the problem of imputation can be found in Kant’s distinction between the Wille and the Willkür: “We have distinguished two meanings of will: Wille as practical reason, the legislative function, and Willkür, as the executive faculty of man” (Beck 1960, p. 202). This is because the concept of a free will in Kant can be ambiguous. It can mean a free Wille, that is, the freedom of practical reason. But it can also mean a free Willkür, that is, the free executive function of man. And the freedom that Kant ascribes to a free Wille is not the same as the freedom that Kant ascribes to a free Willkür, as Beck notes: “The freedom of the former [Wille] is autonomy; it gives a law to Willkür. This law is determined by the nature of the Wille and not by anything else in the world, including human nature or the will of God” (Beck 1960, p. 202). This can be contrasted to the freedom of the Willkür: “The freedom of the latter [Willkür] is spontaneity, the faculty of initiating a causal series in nature” (Beck 1960, p. 203). The problem of imputation arises due to the misleading way in which freedom is talked about in some of Kant’s earlier writings. Freedom of the will is treated as a singular thing, and this obscures the distinction between having a free Wille and having a free Willkür, and thereby leads to the unhappy conclusion that we are morally responsible for our moral actions, but not for our immoral actions. But the problem goes away if the expression “freedom” could somehow be taken in two different senses, as either freedom of the Wille or freedom of the Willkür. Then we can see that Kant’s insistence that only moral actions can be considered free is to be understood as a statement about autonomous actions, which does not imply that we are only morally responsible for our moral actions, and not for our immoral ones.
The reason being that moral responsibility for an action is imputed to us because they are free in the sense of spontaneity.\textsuperscript{22}

\textbf{Conclusion}

This chapter deals with freedom and its relation to moral evil. In particular, it was concerned with two types of Kantian freedom – the spontaneity of the free \textit{Willkür} and the autonomy of the free \textit{Wille}. The spontaneity of the free \textit{Willkür} relates to moral evil in the following way: the incorporation thesis states that an incentive has to be incorporated into a maxim in order to determine the power of choice to action. This act of incorporation is spontaneous, and so the \textit{Willkür} freely determines the way in which the moral incentive and the nonmoral incentive are related to each other in the maxim. Insofar as the moral bent of the maxim is determined by whether we had subordinated the moral incentive or the nonmoral one, it follows that this decision was free in the sense necessary for moral responsibility. The autonomy of the free \textit{Wille} relates to moral evil in the following way: it does not. Autonomy is the “property of the will by which it is a law to itself (independently of any property of the objects of volition)” (4:440). The law that is legislated by such a will is always the moral law, and the maxims of a will under the moral law will always be morally good. Heteronomy is what happens “if the will seeks the law that is to

\textsuperscript{22} This is an imprecise formulation of the freedom of spontaneity. Kant does not consider freedom to consist in the power to do otherwise \textit{per se}, as evidenced by the following passage: “But freedom of choice cannot be defined — as some have tried to define it — as the ability to make a choice for or against the law (\textit{libertas indifferentiae}), even though choice as a \textit{phomenon} provides frequent examples of this in experience. For we know freedom (as it first becomes manifest to us through the moral law) only as a \textit{negative} property in us, namely that of not being necessitated to act through any sensible determining grounds” (6:226). It might be said that this negative property of not being necessitated to act through any sensible determining grounds entails the positive property of being able to act for or against the law, but this entailment would only hold for phenomena. For Kant considers God to have the negative property of not being necessitated to act through any sensible determining grounds, but not the positive property of being able to act for or against the law, because God cannot act on the basis of any sensible determining grounds whatsoever. Hence Wood explains freedom of spontaneity in Kant in the following way: “For Kant, freedom consists in the ability to act autonomously even when we do not, but it does not consist in the possibility of acting heteronomously, even if this possibility always does exist for us” (Wood 1984, pp. 80-81).
determine it *anywhere else* than in the fitness of its own maxims for its own giving of universal law” (4:441). A will that seeks the law that is to determine it in its objects would be heteronomous, and only the maxims of such a will can be morally evil. From this, it follows that moral evil is free in the sense of spontaneity, but it is not free in the sense of autonomy. And this allows Kant to escape the problem of imputation, because the rational agent who does a moral evil is free in the sense needed for moral responsibility and moral imputation, despite not being perfectly rational (insofar as a being that is perfectly guided by reason is, according to Kant, incapable of moral evil). Now radical evil is also a moral evil, even though Kant describes it as the formal ground of every deed contrary to law that is given in time (6:31). Therefore, if radical evil is to be ascribed to us as a *moral* evil (rather than a physical evil), then it is necessary that we are free in the sense needed for moral responsibility and moral imputation – we would need to have freedom of the Willkür in the form of spontaneity. And if radical evil is to be ascribed to us as a moral *evil*, then it is necessary that we are unfree in the sense needed to be a heteronomous will, because the will that chooses radical evil is unfree in that it “does not give itself the law; instead the object, by means of its relation to the will, gives the law to it” (4:441).
CHAPTER 2. TWO CONCEPTIONS OF RADICAL EVIL

The tears began to flow and sobs shook him. He gave himself up to them now for the first time on the island; great, shuddering spasms of grief that seemed to wrench his whole body. His voice rose under the black smoke before the burning wreckage of the island; and infected by that emotion, the other little boys began to shake and sob too. And in the middle of them, with filthy body, matted hair, and unwiped nose, Ralph wept for the end of innocence, the darkness of man’s heart...

William Golding, *Lord of the Flies.*

**Introduction**

What is radical evil? This is the most basic question we can ask on the matter. But the answer is not that straightforward. There are places in the *Religion* where Kant describes radical evil in terms of a *propensity*: “By *propensity* (*propensio*) I understand the subjective ground of the possibility of an inclination (habitual desire, *concupiscientia*), insofar as this possibility is contingent for humanity in general” (6:29). And there are other places in the *Religion* where he describes it in terms of a *fundamental maxim*: “The disposition, i.e. the first subjective ground of the adoption of maxims, can only be a single one, and it applies to the entire use of freedom universally. This disposition too, however, must be adopted through the free power of choice, for otherwise it could not be imputed” (6:25). It is not the least bit clear whether the conception of radical evil as a propensity is supposed to be *identical to* or *distinct from* the conception of radical evil as a fundamental maxim. There are authors who follow the language of the *Religion* and identify the two: “Construed in dispositional terms as a deliberative tendency to rank in a certain way the basic incentives of morality and self-love in their incorporation into first-order maxims, the meta-maxim may be viewed as a propensity. Since evil has already been located in the subordination within a maxim of moral requirements to those of self-love, it follows that by a propensity to evil must be understood the meta-maxim to order the incentives in just this way in
the adoption of an agent’s first order maxims” (Allison 2002, p. 340). But there are others who argue that the two are distinct, because the fundamental maxim belongs to the noumenal world and the propensity belongs to the phenomenal world: “Gesinnung is a constant, an atemporal orientation providing us with the fundamental model for the regulation of our particular temporal choices. So even were we to take Hang to be a technical term badly translated by the English word ‘propensity’, there still seems to be no conceptual sphere for any distinction between having such a Hang, however we might conceive it, and simply being evil, if having it is taken to be equivalent to possession of an evil disposition” (Morgan 2005, p. 97).

This essay comes down on the side of those who consider this conception of radical evil as a propensity and the conception of radical evil as a fundamental maxim to be distinct. For if we compare the basis for positing radical evil as a propensity, we will find them to be very different from the basis for positing radical evil as a fundamental maxim. The basis for positing radical evil as a fundamental maxim lies in the nature of maxims. By their nature, maxims are able to be arranged hierarchically, and the maxim that is positioned at the apex of the hierarchy is the fundamental maxim. If radical evil is to be the first subjective ground of the adoption of maxims, it would have to be an evil fundamental maxim. By contrast, the basis for positing radical evil as a propensity lies in the nature of a disposition. It is needed to explain the existence of moral evil in the world, because moral evil is a possibility of every human being, and a propensity to evil is needed to ground this possibility for moral evil in every human being. But the distinction between the conception of radical evil as a propensity and the conception of radical evil as a fundamental maxim is most evident when it comes to the discussion of the gradations of evil and their grounding in the evil of human nature. Insofar as the two conceptions of radical evil have not been distinctly separated by Kant in his writing, the role it is supposed to play is the grounding of frailty and
depravity is ambiguous. For frailty has as its ground a subjective ground for the possibility of an inclination. The relevant concept of radical evil with respect to frailty is the conception of radical evil as a propensity. By contrast, depravity has as its ground the first subjective ground of the adoption of maxims. The evil of depravity consists not in an inclination but a maxim. The relevant concept of radical evil with respect to depravity is the conception of radical evil as a fundamental maxim. Insofar as the distinction between frailty and depravity collapses if the distinction between the conception of radical evil as a propensity and the conception of radical evil as a fundamental maxim is not maintained, we must be careful to keep the two conceptions of radical evil separate.

Radical Evil as Fundamental Maxim

Radical evil has to be free – but is the freedom that Kant assigns to radical evil the freedom of the Wille or the freedom of the Willkür? From our earlier deliberations, it seems that radical evil cannot be free in the sense of a free Wille. The Wille is the legislative aspect of the will. In this capacity, it prescribes practical principles according to which the human being chooses to act. These practical principles can either be pure, or they can be empirical. For Kant, a Wille that prescribes pure practical principles would be a law unto itself. Insofar as it is a law unto itself, it would be a self-determining Wille, and this Wille would therefore be free in the sense of autonomy. By contrast, a Wille that turns to outside sources would prescribe practical principles that are empirical. This would not be a self-determining Wille, and it would therefore be unfree in the sense

---

23 In his discussion, Allison explains the legislative function of the Wille in the following way: “... To begin with, all of the formulations agree in equating Wille, or will in its legislative function, with practical reason. Considered as such, Wille is the source of laws that confront the human Willkür as imperatives. Although Kant is silent on the point, it seems clear that this must include both the categorical and hypothetical imperatives or, more generally, moral and prudential principles. Both are higher-order rules for governing our selection of maxims and both are products of practical reason. Both, therefore, must be attributed to Wille. Correlatively, it is Willkür, or will in its executive function, that can be said to act, that is, to decide, choose, and even wish under the governance of Wille” (Allison 1990, p. 130).
that Kant calls *heteronomy*. Now radical evil cannot be represented by a pure practical principle, otherwise we would have to say that *radical evil is not evil*, which would be absurd. But an empirical practical principle is always conditioned upon something else, and this implies that the *Wille* must turn to outside sources in its construction of the practical principle. Therefore, the freedom ascribed to radical evil cannot be freedom in the sense of autonomy. This leaves freedom in the sense of a free *Willkür*. The *Willkür* is the executive aspect of the will. In this capacity, it chooses between the practical principles that the *Wille* prescribes to it. Freedom in the sense of the *Willkür* is therefore pertinent to the ascription of moral responsibility, because it is the *Willkür* that chooses and is responsible for the practical principles that we adopt. Now radical evil is considered free for the purposes of moral responsibility: “… But otherwise the use or abuse of the human being’s power of choice with respect to the moral law could not be imputed to him, nor could the good or evil in him be called “moral”)” (6:21). And this shows that the freedom ascribed to radical evil has to do with the freedom of the *Willkür*.

Because radical evil is freely chosen by the *Willkür*, it follows that radical evil has to take the form of a maxim. This is because the *Willkür* makes its choices by subordinating either the moral incentive or the nonmoral incentive that is present in the maxim. And it is therefore up to us whether to subordinate the moral incentive or the nonmoral incentive in our maxim. Because this is a choice that we have to make in order for any action to be possible, it is reasonable to ask for the grounds upon which this selection was made.\(^\text{24}\) Kant believes that there are only two candidates

\(^{24}\) In the Preface, Kant explains this in the following manner: “Yet it is one of the inescapable limitations of human beings and of their practical faculty of reason (perhaps of that faculty in all other worldly beings as well) to be concerned in every action with its result, seeking something in it that might serve them as an end and even prove the purity of their intention – which result would indeed come last in practice (*nexu effectivo*) but first in representation and intention (*nexu finali*)” (6:7n). Kant’s point in this passage is that, unless we represent to ourselves the end of the action, the action itself cannot be represented, insofar as the action is the means by which we accomplish this end. And once we have represented to ourselves an end towards which our actions may be directed, the existence of this representation and intention provides us with an *incentive* to action.
available: either the selection was made on the basis of a maxim or it was made on the basis of a natural impulse. If the selection of the maxim was made on the basis of a natural impulse, then it could not have been free. But if the selection of the maxim was made on the basis of another maxim, then this selection would involve a choice between the moral incentive and the nonmoral incentive. This would in turn allow us to ask once again for the grounds upon which the selection was made. If that selection was made on the basis of a natural impulse, however, then the entire series of grounds subsequent to it would have been the result of the determination of natural causes, in which case the evil deed would not be imputable to us because it would not have been free. In the series of grounds leading up to an imputable act, none of these grounds can turn out to be a natural impulse. For otherwise, we would have a deed that is performed from a sensory inducement, which is a contradiction (6:31). Insofar as radical evil must be a part of this series of grounds, and indeed the origin of this series, if it should turn out that radical evil was not chosen through a fundamental maxim, then not one member of this entire series could possibly be the product of the absolute spontaneity of the power of choice. This would mean, whatever the deed that results, it would not be a deed that can be imputable to us. Thus, whichever the incentive we choose to subordinate in the maxim, this choice must be made on the basis of a practical principle, not on the basis of natural impulses. And this means that the Willkür freely chooses the ordering of the moral incentives and the nonmoral ones in the maxim.

This idea that there is an entire series of grounds for our free actions is prominent in Allison’s discussion of maxims. For Allison, maxims are arranged in a hierarchical order: “… One might think of maxims, in analogy with concepts (considered intensionally), as arranged hierarchically, with the more general embedded in the more specific, like genera in species” (Allison 1990, p. 93). We can see why this is so, given a proper understanding of what maxims are
for Kant. The maxim is defined in the *Groundwork of the Metaphysics of Morals* as the subjective principle of acting (4:421n). It contains the incentives that move the subject to action, be they moral incentives or nonmoral ones, and once the subject has incorporated them into his maxim, the maxim becomes a universal rule for the subject, according to which the subject wills to conduct himself (6:24). These incentives are the grounds that the subject is motivated to act on. If these incentives were not present, it would be impossible for any reference to an end to be made (and in the absence of all reference to an end no determination of the will can take place in human beings at all) (6:4). And once we see that maxims are principles that contain the grounds for an action in them, it becomes easier to understand how our maxims are supposed to relate to each other. For the grounds that we have for acting can have different degrees of generality, as Allison writes: “… As rules dictating action types rather than particular actions, maxims, like concepts, are general with respect to the number of possible items that fall under them. Accordingly, there are always (in principle at least) a number of distinct ways in which an agent can act upon a maxim, just as there are a number of different ways in which one can attempt to realize an end that one has chosen for oneself” (Allison 1990, p. 90). Therefore, the hierarchy of maxims is formed based on the grounds that are contained in them. The more general the ground contained in a maxim, the higher its place in the hierarchy, and the more maxims (with less general grounds) that fall under it.

Let us put this discussion in more concrete terms. For example, consider a student who is reading a book in order to prepare for a test the next day. The desire to do well in that test *motivates* him to read the book, and it constitutes the material for a maxim which he makes into a universal rule for himself, according to which he wills to conduct himself. But the student might also be reading the book in preparation for a future career as a lawyer. And the desire for fame and fortune also *motivates* him to read the book, and it constitutes the material for a maxim which he makes
into a universal rule for himself, according to which he wills to conduct himself. There is no reason why both these grounds cannot be operative on the student at the same time. Each of these grounds motivate the student to read the book. Their difference consists mainly in this: The desire to do well in the test motivates the student to read the book, and little else. By contrast, the desire for fame and fortune motivates the student to read the book, but it also motivates him to a great many other things besides. In other words, the student's desire for fame and fortune explains his desire to do well in the test. This is what Caswell had in mind when he observed that maxims can be general to different degrees:

Note that maxims always express a certain level of generality. For while non-free, irrational beings such as animals might determine their behavior according to particular representations, rational agency is a power of determining desire through concepts, or general discursive representations. This means that the same maxim can govern a number of actions that take place in different times and places, but it also means that maxims, like concepts, can be fit under another maxim's broader 'extension'. Thus, an agent's maxims form a hierarchy, with more particular principles fitting under the more general. Indeed, it must be the case that each maxim has its place within a hierarchical system of maxims, for its location in that season provides the rational constraints which constitute the formal conditions of willing the maxim (Caswell 2006, p. 193).

The fact that an agent's motives fit together in this way gives us the reason as to why there needs to be a highest place, or fundamental maxim, and the reason is this: What it means for the adoption of a maxim to be grounded in a higher maxim is, most straightforwardly, that the maxim derives its practical justification from the higher maxim. This reason is taken from a parallel case in the field of epistemology. For beliefs, too, form a similar hierarchy, in which beliefs are held on the basis of their grounding in other beliefs. Here we might say: What it means for a belief to be grounded in another belief is, most straightforwardly, that the belief is inferentially justified by the other belief. From this, we are able to construct the regress argument for epistemic foundationalism, which Aristotle first proposed in the Posterior Analytics (PA 72b5-35). For if we
hold on to a belief A because it is inferentially justified by the belief B, then the belief B had better be a belief we hold, and are justified in believing. After all, if we are not justified in believing B, then it seems impossible for the belief A to be inferentially justified by the belief B. But how would we be justified in believing B? If we say that we believe B because it is inferentially justified by the belief C, then an infinite regress is in the making, because we will soon be forced to posit a belief D to inferentially justify the belief C, and so on and so forth. In order to avoid an infinite regress, Aristotle believed that we would need to accept epistemic foundationalism. That is, Aristotle believed that there needed to be beliefs that are basic—beliefs that are able to justify other beliefs and are not themselves justified by other beliefs.

Returning to the discussion at hand, we may say that when a lower maxim derives its practical justification from the higher maxim, this means that the higher maxim explains why the agent has the lower maxim that he has—and it does so by reference to the motives that constitute the material of the respective maxims. This higher maxim must itself be grounded in a still higher

---

25 This conclusion skips a few steps. Aristotle’s is a disjunctive argument, in which he considers the different positions that may be taken with respect to this infinite regress of beliefs. Then he eliminates the other positions until he is left with the position of epistemic foundationalism. First, we might simply accept that the regress is infinite. Aristotle describes this position as skepticism when, after accepting that the regress is infinite, the conclusion that is drawn is that there are no justified beliefs whatsoever: “The first school, assuming that there is no way of knowing other than by demonstration, maintain that an infinite regress is involved, on the ground that if behind the prior stands no primary, we could not know the posterior through the prior (wherein they are right, for one cannot traverse an infinite series)” (PA 72b8-12). Second, we might accept that every belief needs to be inferentially justified, but deny that this entails an infinite regress. Instead, beliefs become justified insofar as they inferentially justify each other, a position known as coherentism. This solution to the regress problem was rejected by Aristotle because he considered it to be circular reasoning: “Now demonstration must be based on premisses prior to and better known than the conclusion; and the same thing cannot simultaneously be both prior and posterior to one another: so circular demonstration is clearly not possible in the unqualified sense of demonstration” (PA 72b25-30). Third, we might accept that there are basic beliefs, that is, beliefs that can inferentially justify other beliefs without being justified by inference themselves. Insofar as the regress of inferential justifiers halts at these beliefs, the regress of inferential justifiers never goes infinite, and hence there is no regress problem. This is the position known as foundationalism, as Aristotle writes: “Our own doctrine is that not all knowledge is demonstrative: on the contrary, know of the immediate premisses is independent of demonstration. (The necessity of this is obvious; for since we must known the prior premisses from which the demonstration is drawn, and since the regress must end in immediate truths, those truths must be indemonstrable)” (PA 72b18-24). By process of elimination, Aristotle and other defenders of epistemic foundationalism argue that the only tenable position is that of foundationalism, hence the regress problem is oftentimes considered an argument for foundationalism.
maxim, however, otherwise the maxim would not have any practical justification, and the lower maxims that fall under this higher maxim would also be deprived of their practical justification, insofar as their being practically justified presupposes a source that is itself practically justified. Because this process can be repeated at every stage of the analysis, we are faced with a regress of practical justifiers, and this puts us in a position similar to what we described in the previous paragraph, in the parallel case of epistemic justification. The regress of practical justifiers is halted in much the same way. That is, we must admit that the regress of practical justifiers has to come to an end at a certain point in order to avoid an infinite regress, and this entails that there must be a highest maxim that grounds the adoption of all other maxims: “… But there cannot be any further cognition of the subjective ground or the cause of this adoption (although we cannot avoid asking about it), for otherwise we would have to adduce still another maxim into which this disposition would have to be incorporated, and this maxim must in turn have its ground” (6:25).

26 Following Aristotle, we might therefore proceed using a disjunctive argument. In other words, it must be the case that either this regress of practical justifiers is endless or that this regress of practical justifiers circles back upon itself or that this regress of practical justifiers ends at a certain point, that is, the same three options from the parallel case of epistemic justification. The first option, where we admit that the regress of practical justifiers is endless, entails that for every maxim that is practically justified, there is a higher maxim that practically justifies it, and this process proceeds unto infinity. The problem is that this solution does not seem to reflect the realities of the way we reason, since we do not have access to an infinite chain of grounds, and we would not be able to provide such an infinite chain if asked, given our finite natures. The second option, where we admit that the regress of practical justifiers circles back upon itself, entails that a maxim ultimately derives the practical justification it has from itself, but that this circularity is acceptable so long as the circle of practical justification is large enough. The problem is that this solution is an endorsement of circular reasoning, because no matter how large the circle of practical justifications might be, a maxim that needs to be grounded by another maxim cannot, by this logic, ground itself. Therefore, only the third option remains. This would be the foundationalist position detailed earlier. But instead of basic beliefs that justify other beliefs without needing another belief to justify it, we have here a practical justifier that is basic. That is, a practical justifier that justifies other practical justifiers without needing another practical justifier to justify it. This could be a belief, but in Kant’s case, the practical justifier is a maxim, which would be a practical principle and not a belief.

27 Regarding the evil fundamental maxim, Kant writes: “This evil is radical, since it corrupts the ground of all maxims; as natural propensity, it is also not to be extirpated through human forces; for this could only happen through good maxims – something that cannot take place if the subjective supreme ground of all maxims is presupposed to be corrupted” (6:37). This follows from the discussion about the hierarchical nature of maxims. The presence of a lower maxim in an agent’s hierarchy of maxims is explained by reference to higher maxims in that hierarchy. If, then, there is to be a change to a lower maxim in an agent’s hierarchy of maxims, this change is also to be explained by reference to higher maxims in that hierarchy. Insofar as the fundamental maxim is the highest maxim there is, there is no higher maxim above it, and therefore no maxim that is capable of bringing about a change of the fundamental maxim. In order to change from an evil fundamental maxim to a good fundamental maxim, divine intervention would be needed. This requires the type of grace that María is describing in the following passage: “An explanation of the very change
The regress argument proves that the hierarchy of maxims has to be grounded in a fundamental maxim. And this fundamental maxim has to be either good or evil. This is due to the position known as ethical rigorism. According to Kant, ethical rigorism is opposed by the position known as ethical latitudinarianism, which comes in two forms: indifferentism and syncretism. The former portrays human actions (and nature) as neither good nor evil, whereas the latter portrays human actions (and nature) as partly good and partly evil. Even though Kant admits that experience points to a middle position between the two extremes of good and evil, and therefore to ethical latitudinarianism, he goes on to write: “It is of great consequence to ethics in general, however, to preclude so far as possible, anything morally intermediate, either in actions (adiaphora) or in human character; for with any such ambiguity all maxims run the risk of losing their determination and stability” (6:22). This is because maxims are the subjective principles of an action (4:401).

in the fundamental disposition itself in terms of divine activity implies that God’s action affects the most fundamental desires of the person, including the person’s desires about the kinds of desires he or she wants to have. This means nothing less than God changing the whole character of the person. Kant does not deny that God may have such an effect on persons, and he even tells us that this kind of action on God’s part may, for all we know, be able to coexist side by side with freedom. Yet he correctly insists that we cannot make any practical use of such an idea. We cannot, because this kind of action of God bypasses the personhood of the agent” (Mariña 1997, pp. 386-387).

There is an issue with the account of adiaphora that Kant has presented here. In the Religion, Kant brings up adiaphora in his discussion of ethical rigorism, where he writes: “It is of great consequence to ethics in general, however, to preclude, so far as possible, anything morally intermediate, either in actions (adiaphora) or in human characters; for with any such ambiguity all maxims run the risk of losing their determination and stability. Those who adhere to this strict way of thinking are commonly called rigorists (a name intended to carry reproach, but in fact a praise); so we can call latitudinarians those at the opposite extreme” (6:22). Insofar as Kant identifies himself as an ethical rigorist, it stands to reason that he intends to preclude anything morally intermediate, in order to avoid causing the maxims to lose their determination and stability. Therefore, Kant must preclude moral intermediacy in actions and in human characters. This seems to be inconsistent with the account of adiaphora given in the Doctrine of Virtue, where adiaphora is mentioned in Kant’s criticism of “fantastic virtue”: “… But that the human being can be called fantastically virtuous who allows nothing to be morally indifferent (adiaphora) and strews all his steps with duties, as with mantraps; it is not indifferent to him whether I eat meat or fish, drink beer or wine, supposing that both agree with me. Fantastic virtue is a concern with petty details which, were it admitted into the doctrine of virtue, would turn the government of virtue into tyranny” (6:409). Here Kant chides the fantastically virtuous for allowing nothing to be morally indifferent. For there are actions that are morally indifferent – it does not matter whether I eat meat or eat fish, supposing both agree with me. Therefore, the fantastically virtuous are misguided in thinking that it makes a moral difference whether I eat meat or I eat fish. These actions are morally indifferent, and are called adiaphora in the Metaphysics of Morals.

There is a way to resolve this inconsistency. It is to acknowledge that Kant uses adiaphora in the two passages without making a distinction between adiaphora as something that is morally intermediate in the Religion and adiaphora as something that is morally indifferent in the Doctrine of Virtue. In order to explain this distinction more clearly, consider it in the context of the following passage from the Metaphysics of Morals: “… If the law can prescribe
In order to determine the deontic status of an action, we are supposed to evaluate the action’s maxim in light of the categorical imperative – an action is permissible when its maxim can be willed as a universal law and is forbidden if its maxim cannot be willed as a universal law (4:402). But if ethical latitudinarianism is true, then it would be possible for there to be actions that are neither good nor evil, or for there to be actions that are partly good and partly evil. Insofar as the basis of these judgments can only be made by evaluating the action’s maxims, it follows from ethical latitudinarianism that it would have to be possible for the maxims themselves to be neither good nor evil, or for the maxims themselves to be partly good and partly evil, and the determination and stability of the maxims would be lost.

Kant appeals to the incorporation thesis in order to disprove this result. The incorporation thesis consists in the following statement: “… Freedom of the power of choice has the characteristic, entirely peculiar to it, that it cannot be determined to action through any incentive except so far as the human being has incorporated it into his maxim (has made it into a universal rule for himself, according to which he wills to conduct himself)” (6:24). Now there are two types of ethical latitudinarianism, according to Kant. Indifferentism is the species of ethical latitudinarianism in which actions are neither good nor evil (morally neutral). Whereas syncretism

only the maxim of actions, not actions themselves, this is a sign that it leaves a playroom (latitudo) for free choice in following (complying with) the law, that is, that the law cannot specify precisely in what way one is to act and how much one is to do by the action for an end that is also a duty. – But a wide duty is not to be taken as permission to make exceptions to the maxim of actions but only as permission to limit one maxim of duty by another (e.g., love of one’s neighbor in general by love of one’s parents), by which in fact the field for the practice of virtue is widened” (6:390). Kant makes a distinction in his philosophy between perfect and imperfect duties, where perfect duties are duties which admit of no exceptions, and imperfect duties are duties that can be fulfilled in a multiplicity of different ways. If the fantastically virtuous attempts to dictate the way in which I am to pursue my imperfect duties (e.g. in my duty to take care of my animal needs, he dictates that I eat meat, and not fish, even though both agree with me), then this would turn the “government of virtue into tyranny” (6:409). For it is morally indifferent how I fulfill my imperfect duties, so long as I do perform them. This does not mean, however, that these acts that are morally indifferent are also morally intermediate. It does not matter whether I eat meat or I eat fish, as long as both agree with me, because doing either would be a fulfillment of my imperfect duty. But in fulfilling an imperfect duty, my actions are not morally intermediate, because I am still fulfilling a duty of mine. And if this is so, then there is no danger of ambiguity in this, and therefore no risk of the maxims “losing their determination and stability” (6:22).
is the species of ethical latitudinarianism in which actions are partly good and partly evil (of mixed morality). But the maxim of an action that is neither good nor evil would have to itself be neither good nor evil, and the maxim of an action that is partly good and partly evil would have to itself be partly good and partly evil. This is impossible. For a maxim is good or evil depending on the way the moral incentive or the nonmoral incentive has been incorporated into the maxim. If the nonmoral incentive is subordinated to the moral incentive during the process of incorporation, then the maxim is good. And if the moral incentive is subordinated to the nonmoral incentive during the process of incorporation, then the maxim is evil.

For the action to turn out neither good nor evil, as the indifferentist asserts, the human being would have had to avoid subordinating either the moral incentive or the nonmoral incentive when incorporating them into his maxim, which is impossible. As Kant writes: “In us, however, the law is incentive, = a. Hence the lack of the agreement of the power of choice with it (= 0) is possible only as the consequence of a real and opposite determination of the power of choice, i.e. of a resistance on its part, = –a; or again, it is only possible through an evil power of choice” (6:22n). There can be no such thing as an action that is neither good nor evil, because an action that is not determined by the moral law would, by definition, be evil. It entails the elevation of the nonmoral incentive over the moral incentive in the maxim. And this results in a maxim that is evil. Likewise, for the action to turn out partly good and partly evil, as the syncretist asserts, the human being would have to subordinate both the moral and the nonmoral incentive when incorporating them into his maxim, which is impossible. For Kant writes: “Nor can a human being be morally good in some parts, and at the same time evil in others. For if he is good in one part, he has incorporated the moral law into his maxim, and were he, therefore, to be evil, in some other part, since the moral law of compliance with duty in general is a single one and universal, the maxim
relating to it would be universal yet particular at the same time: which is contradictory” (6:24-25). This reasoning applies to actions as well. If an action is good, this means the human being has subordinated the nonmoral incentive to the moral incentive during the process of incorporation, resulting in a good maxim. And if an action is evil, this means the human being has subordinated the moral incentive to the nonmoral incentive during the process of incorporation, resulting in an evil maxim. For an action to be partly good and partly evil, then, both of these must be true: the human being has subordinated the nonmoral incentive to the moral incentive and he has subordinated the moral incentive to the nonmoral incentive in the same maxim. And this is absurd.

Ethical latitudinarianism is therefore an untenable philosophical position, and this fact tells against ethical latitudinarianism and for ethical rigorism.

And Kant extends this argument for ethical rigorism to cover human nature or human character also. If ethical latitudinarianism is true, then it would be possible for a person’s character to be neither good nor evil (indifferentism), or for his character to be partly good and partly evil (syncretism). And the person’s fundamental maxim would have to reflect this fact: he would either have a fundamental maxim that is neither good nor evil, or he would have a fundamental maxim that is partly good and partly evil. Yet the same considerations as before are applicable to the fundamental maxim also. The fundamental maxim is either good or evil depending on whether the moral incentive is subordinated to the nonmoral incentive during the process of incorporation by the Willkür, or vice versa. But a fundamental maxim cannot be morally indifferent (neither good nor evil), as explained previously. For there can be no such thing as a character that is neither good nor evil, because a character that is not determined by the moral law is, by definition, evil. It entails the elevation of the nonmoral incentive over the moral incentive in the fundamental maxim. And this results in a fundamental maxim that is evil. Nor can a fundamental maxim be of mixed morality
(partly good and partly evil). For this would mean that the human being has subordinated the nonmoral incentive to the moral incentive and he has subordinated the moral incentive to the nonmoral incentive in the same fundamental maxim. And this is absurd. For, as Allison has argued, a person’s character and his fundamental maxim are one and the same: “If one is to avoid regarding an agent’s Gesinnung as a product of nature rather than freedom, which would effectively undermine morality by reducing goodness and badness of will to a matter of constitutive moral luck, then it is necessary to consider it as itself chosen. But as chosen, this Gesinnung must itself be conceived as a maxim, albeit one of the highest order, that provides a direction or orientation for the moral life of the agent viewed as a whole” (Allison 1990, p. 141). And this would mean that a person’s character could no more be neither good nor evil, or be partly good and partly evil, than a person’s fundamental maxim be neither good nor evil, or be partly good and partly evil. And this fact tells against ethical latitudinarianism and for ethical rigorism.

But even accepting Kant’s ethical rigorism, what reason do we have for assuming that this fundamental maxim is evil rather than good? After all, experience tells us that every human being acts from both good maxims and bad ones, and there seems to be no reason to assume that, as we traverse the hierarchy of maxims, the maxim at the apex of this hierarchy would be evil, rather than vice versa. For consistency’s sake, it seems that, if we assume that evil maxims must ultimately be grounded in a fundamental maxim that is evil, then we must also assume that good maxims must ultimately be grounded in a fundamental maxim that is good. But Kant rejects this line of reasoning. This is because the categorical imperative consists in the following command: act only in accordance with that maxim through which you can at the same time will that it become a universal law. If the hierarchy of maxims is grounded in a fundamental maxim that is good, we would see a consistent application of the categorical imperative to the maxims within the hierarchy.
Every maxim in the hierarchy would be determined to have the form of universal law. But if the hierarchy is grounded in a fundamental maxim that is evil, we would not see the consistent application of the categorical imperative to the maxims within the hierarchy. There will be some maxims that are determined to have the form of universal law, and others that are not. This leads Allison to observe that the evil fundamental maxim consists in an inconsistency: “… It can be more properly described as the fundamental or ruling maxim to license exceptions to one’s own case to moral requirements (“moral holidays”, as it were) while still acknowledging the general validity of these requirements” (Allison 1990, p. 151). The asymmetry between the two cases, wherein anything less than the consistent application of the categorical imperative to one’s hierarchy of maxims is a sign that the fundamental maxim at the apex of this hierarchy has been corrupted by evil, is at the root of Kant’s doctrine of radical evil. For it is by appeal to this criteria (that anything less than the consistent application of the categorical imperative to the maxims in one’s hierarchy is a sign that the fundamental maxim at the apex of this hierarchy has been corrupted by evil) that Kant judges human nature to be evil, because all have fallen short when it comes to complete compliance with duty in general.

The Anthropological Evidence for Radical Evil

But how does Kant defend his assertion that every human being has fallen short of complete compliance with duty in general? There is a point in his discussion of radical evil where Kant seems like he is about to prove to us his assertion that human beings are by nature evil. But then he abruptly stops and writes this instead: “We can spare ourselves the formal proof that there must be such a corrupt propensity rooted in the human being, in view of the woeful examples that the experience of human deeds parades before us” (6:32). The first set of these woeful examples that the experience of human deeds parades before us is the savagery exhibited by the human beings
in the so-called state of nature. And here Kant describes the cruelty of the tribesmen of New Zealand in their ritual murders and the wanton slaughter that the Arathapescow and Dog Rib Indians of northwestern America engage in. The second set of these woeful examples that the experience of human deeds parades before us is the ressentiment that civilization breeds in human beings. And so Kant describes the way prudence prescribes mistrust between even the closest friends, the propensity to hate those we are indebted to, and the schadenfreude that we secretly derive from the misfortune of others. The third set of these woeful examples that the experience of human deeds parades before us is the behavior of nation-states in their international affairs. And this prompts Kant to bemoan the state of constant war between nation-states that makes the philosophical chiliasm of a perpetual peace an absurd fantasy. These three sets of woeful examples that the example of human deeds parades before us are supposed to prove the universal existence of the propensity to evil in human beings. As such, a natural way of interpreting Kant’s argument here is to construe it as an empirical generalization. Thus, we find Wood writing in Kant’s Moral Religion: “When Kant says that man is evil “by nature”, he does not mean to explain evil, but only to point out the universality of evil in man. Kant thus looks for evidence supporting the claim that all men, without exception, exhibit a propensity to evil; and he finds such evidence in “the multitude of crying examples which the experience of the actions of men puts before our eyes” (Wood 1970, p. 225).

---

29 Wood’s position on Kant’s argument for radical evil changed drastically between the publication of Kant’s Moral Religion (1970) and the publication of Kant’s Ethical Thought (1999). In Kant’s Moral Religion, Wood interpreted Kant’s argument for radical evil as an empirical generalization from the anthropological evidence that was provided. The multitude of woeful examples that Kant provides are meant to show that “in spite of the differing conditions in which we find man, a propensity to evil, to lie, to kill his fellows, or to enslave and exploit them, to adopt any course of action which leads to the satisfaction of personal wishes, is always characteristic of him” (Wood 1970, p. 226). By the time of Kant’s Ethical Thought, Wood had turned to the interpretation put forth by Sharon Anderson-Gold: “… The doctrine of radical evil is based on Kantian anthropology, that is, on his theory of the purposive development of the human race’s collective predispositions, which nature has brought about through the human trait of unsociable sociability. This makes radical evil an empirical thesis without reducing it to a mere inductive generalization. On this interpretation, radical evil would pertain to us insofar as we are social beings; the
This solution has drawn the ire of almost every interpreter that has written on the topic of radical evil. Their frustration with Kant is understandable. As Morgan observes: “... Lacking as he obviously does empirical acquaintance with the behavior of each and every human being past, present and future, Kant will need to present an \textit{a priori} argument in order to earn the right to assert that all human beings have such a propensity. Indeed, since he thinks that this characteristic is necessary, even if we were \textit{per impossibile} to have such an acquaintance, this would still not underwrite the modality of the claim” (Morgan 2005, p. 62). And Kant seems to be perfectly well aware of this, because when he asserts that human beings are by nature evil, he leaves no doubt as to the epistemological status of this assertion. The assertion that human beings are by nature evil must be \textit{a priori} in nature: “In order, then, to call a human being evil, it must be possible to infer \textit{a priori} from a number of consciously evil actions, or even from a single one, an underlying evil maxim, and, from this, the presence in the subject of a common ground, itself a maxim, of all particular morally evil maxims” (6:20). But this \textit{a priori} assertion cannot be an analytic \textit{a priori} statement about human nature. This is because Kant considers radical evil to be a moral propensity, and this means that radical evil must therefore be the product of our freedom of choice. Hence Kant explains that the assertion that human beings are by nature evil “simply means that being evil applies to him considered in his species, not that this quality may be inferred from the concept of his species (from the concept of a human being in general, for then the quality would be necessary)” (6:33). Taken together, these two facts indicate that Kant’s assertion is a synthetic \textit{a priori} proposition, as opposed to an analytic \textit{a priori} proposition or a synthetic \textit{a posteriori} evil in our nature is closely bound up with our tendencies to compare ourselves with others and compete with them for self-worth” (Wood 1999, p. 287). This led Wood to repudiate his 1970 interpretation of Kant’s argument for radical evil in the following passage: “... The confusion engendered by these remarks [6:32-33] explains the wide disagreement among commentators, whose opinions run all the way from my own naïve conjecture long ago that the doctrine of radical evil is intended simply as an empirical generalization to Henry Allison’s equally implausible proposal that it is not a far-reaching indictment of the human will but only a trivial practical corollary of our finitude” (Wood 1999, p. 287).
proposition. But the multitude of woeful examples that the experience of human deeds parades before us can only ever amount to an *a posteriori* proposition. Hence Morgan’s criticism: “All that the woeful parade of human deeds can show us is that there are evil people, or at most, that evil is common and widespread. It would be an entirely reckless generalization to conclude from the undeniably extensive litany of the crimes that human beings have carried out that every single human being has a propensity to evil, and indeed actually is evil, and that the root of this evil lies in human nature” (Morgan 2005, pp. 62-63). For the examples Kant gives fall far short of what he needs to prove his assertion that human beings by nature evil, and interpreters have been wont to pass over the anthropological discussion of radical evil in near-silence ever since.

In this section, let us delve into the multitude of woeful examples that the experience of human deeds parades before us. The intention here is to provide some context for the three sets of examples that are given, so that we can see why Kant would have wanted to include them in his proof that human beings are by nature evil. First are the evils that occur in the state of nature. Kant writes the following indictment:

If we wish to draw our examples from that state in which many a philosopher especially hoped to meet the natural goodliness of human nature, namely from the so-called *state of nature*, let one but compare with this hypothesis the scenes of unprovoked cruelty in the ritual murders of Tofoa, New Zealand, and the Navigator Islands, and the never-ending cruelty (which Captain Hearne reports) in the wide wastes of northwestern America from which, indeed, no human being derives the least benefit, and we find vices of savagery more than sufficient to distance us from any such opinion (6:33).

The first thing to notice is that Kant had a specific target in mind when he cited these examples of evil happening in the *state of nature*: the philosophers that assert the goodness of the human being in the state of nature. This is the philosophical position of Rousseau, who had written: “Everything good comes from the hand of the author of things; everything degenerates between the hands of man” (Emile 4:35). For Rousseau, the social condition has a corrupting effect on human nature,
and it is the reason why human nature is evil. But he posits a time prior to human society, when human beings lived in the state of nature, and these human beings living in the state of nature had a natural goodliness to them.\textsuperscript{30} It is in response to this hypothesis that the human being is good in the state of nature, and only becomes evil due to the detrimental effects of society, that Kant cites the ritual murders of Tofoa and the never-ending cruelty of the wide wastes of northwestern America. In other words, these woeful examples that the experience of human deeds parades before us are being used by Kant as evidence to disprove the Rosseauian hypothesis about the goodness of the human being in the state of nature.

The next set of examples have to do with the human being in his civilized state. Objecting to the idea that civilization brings out the goodness of the human being, Kant writes the following indictment:

\begin{quote}
If we are however disposed to the opinion that we can have a better cognition of human nature known in its civilized state (where its predispositions can be more fully developed), we must then hear out a long melancholy litany of charges against humankind—of secret falsity even in the most intimate friendship, so that a restraint
\end{quote}

\textsuperscript{30} Rousseau discusses this further in the Discourse on the Origin and Basis of Inequality Among Men (1754). The hypothetical nature of the discussion in the Discourse is evident when he writes: “The investigations we may enter into, in treating this subject, must not be considered as historical truths, but only as mere conditional and hypothetical reasonings, rather calculated to explain the nature of things, than to ascertain their actual origin; just like the hypotheses which our physicists daily form respecting the formation of the world” (Rousseau 1754, pp. 175-176). The Discourse is divided into two parts. Part I of the Discourse is concerned with the original state of man. It is here that Rousseau expounds on his thesis that man was originally good when he was still in the state of nature. Hence he rejects the Hobbesian thesis that man in the state of nature must be naturally wicked: “Hobbes did not reflect that the same cause, which prevents a savage from making use of his reason, as our jurists hold, prevents him also from abusing his faculties, as Hobbes himself allows: so that it may be justly said that savages are not bad merely because they do not know what it is to be good: for it is neither the development of the understanding nor the restraint of law that hinders them from doing ill; but the peacefulness of their passions, and their ignorance of vice” (Rousseau 1754, pp. 196-197). And Part II of the Discourse is concerned with how the wickedness of man’s current state came to be. This Rousseau traces to its origins in the malignant influence of human society: “… I think I have shown man to be naturally good. What then can have deformed him to such an extent, except the advances he has made, and the knowledge he has acquired? We may admire human society as much as we please; it will be none the less true that it necessarily leads men to hate each other in proportion as their interests clash, and to do one another apparent services, while they are really doing every imaginable mischief” (Rousseau 1754, pp. 239-240). Together, these two theses are used by Rousseau to confirm the downward trajectory of the moral development of man. In denying the thesis that man was originally good when he was still in the state of nature, Kant rejects the idea that there is a downward trajectory to the moral development of man. The anthropological evidence he has provided, in the wickedness of both the savage and the civilized man, shows that there has been little development, morally speaking, in the character of the species throughout the ages.
on trust in the mutual confidence of even the best friends is reckoned a universal maxim of prudence in social dealings; of a propensity to hate him to whom we are indebted, to which a benefactor must always heed; of a hearty goodwill that nonetheless admits the remark that “in the misfortunes of our best friends there is something that does not altogether displease us”; and of many other vices yet hidden under the appearance of virtue... (6:33).

Here again, notice that Kant has a specific target in mind when he cites these examples of evil: the philosophers that assert the goodness of the human being in his civilized state. This is the philosophical position held by many of Kant’s Enlightenment contemporaries. Lest we forget, Kant's doctrine of radical evil drew the condemnation it did from his contemporaries precisely because of the way it ran counter to this dogma of Enlightenment thinking, an attitude that Goethe captured in his criticism of the Religion in the following passage: “Kant required a long lifetime to purify his philosophical mantle of many impurities and prejudices. And now he has wantonly tainted it with the shameful strain of radical evil, in order that Christians too might be attracted to kiss its hem” (Fackenheim 1954, p. 340). For Enlightenment thinkers, human life in the state of nature can only be described as “nasty, brutish, and short” (Leviathan 1:13:9). It is only by coming together in civil society that we are able to escape the state of war of all against all (Leviathan 1:13:13). And it is only once we have escaped this state of war of all against all that there is an opportunity for any human goodness to flourish. It is in response to this hypothesis that the

31 Besides Hobbes, there are a number of Enlightenment figures that shared this optimistic perspective on history. In England, Adam Smith’s treatises The Theory of Moral Sentiments (1759) and The Wealth of Nations (1776) featured the idea of an “invisible hand” that guides the economic development of society towards ever-greater production. In so doing, they create societies that are increasingly wealthy, as these advances in the means of production benefit both rich and poor alike: “The rich only select from the heap what is most precious and agreeable. They consume less than the poor, and in spite of their natural selfishness and rapacity, though they mean only their own convenience, though the sole end which they propose from the labors of all the thousands whom they employ, be the gratification of their own vain and insatiable desires, they divide with the poor the produce of all their improvements. They are led by an invisible hand to make nearly the same distribution of the necessitates of life, which would have been made, had the earth been divided into equal portions among all its inhabitants, and thus without intending it, without knowing it, advance the interests of the society, and afford means to the multiplication of the species” (Smith 1759, p. 163). And the French Enlightenment produced Anne-Robert-Jacques Turgot and Nicolas de Condorcet, both of whom believed that the scientific and philosophical developments taking place in Europe would eventually spread across the globe. Thus we find Condorcet, writing in his Outlines of an historical view of the progress of the human mind (1795), declaring:
human being is good in the civilized state that Kant brings up the examples of secret falsity among friends, of *ressentiment*, and of *schadenfreude*. The point Kant is making here is that these evils are no improvement compared to the savage cruelties of the state of nature. If the evils of the state of nature are vices that are grafted onto our predisposition to animality (6:26), these evils are vices of culture that are grafted onto our predisposition to humanity (6:27). Far from being any sort of improvement of the human condition, these vices simply represent the corruption by evil of a different aspect of the human condition. In other words, these woeful examples that the experience of human deeds parades before us are being used by Kant as evidence to *disprove* the Enlightenment hypothesis about the goodness of the human being in the state of civilization.

The last example Kant gives deals with nation-states in their relations to each other. Here is the indictment that Kant writes against them:

So long as a state has a neighboring one which it can hope to subdue, it strives to aggrandize itself by subjugating it. It thus strives for a universal monarchy – a state constitution in which all freedom would necessarily expire, and, together with it, virtue, taste, and science (which follow upon freedom). Yet after this monster (in which the laws gradually lose their force) has swallowed up all its neighbors, it ultimately disintegrates all by itself. It divides through rebellion and factionalism into many smaller states which, instead of striving after a union of states (a republic

… We shall find the strongest reasons to believe, from observation of the progress which the sciences and civilization have hitherto made, and from the analysis of the march of the human understanding, and the development of its faculties, that nature has fixed no limits to our hopes.

If we take a survey of the existing state of the globe, we shall perceive, in the first place, that in Europe the principles of the French constitution are those of every enlightened mind. We shall perceive that they are too widely disseminated, and too openly professed, for the efforts of tyrants and priests to prevent them from penetrating by degrees into the miserable cottages of their slaves, where they will soon revive those embers of good sense, and rouse that silent indignation which the habit of suffering and terror have failed totally to extinguish in the minds of the oppressed.

If we next look at the different nations, we shall observe in each, particular obstacles opposing, or certain dispositions favoring this revolution. We shall distinguish some in which it will be effected, perhaps slowly, by the wisdom of the respective governments; and others in which, rendered violent by resistance, the governments themselves will necessarily be involved in its terrible and rapid motions.

Can it be supposed that either the wisdom or the senseless feuds of European nations, cooperating with the slow but certain effects of the progress of their colonies, will not shortly produce the independence of the entire new world; and that then, European population, lending its aid, will sail to civilize or cause to disappear, even without conquest, those savage nations still occupying there immense tracts of country (Condorcet 1796, pp. 252-254).
of free federated peoples), in turn begin the same game all over again, so that war (that scourge of the human race) will not cease (6:34).

This example differs from the previous ones. It seems to be a continuation of the discussion of the goodness of human beings in their civilized state. Hence Kant writes: “And if we are not satisfied yet, we need but consider a state wondrously compounded from both the others, namely that of a people in its external relations, where civilized peoples stand vis-à-vis one another in the relation of raw nature (the state of constant war) and have also firmly taken it into their hands not to get out of it” (6:34). In this example, Kant is using the state as a macrocosm of the individual human being. The point that he is making here is that the brutality of the state of nature never really

---

Morgenthau writes in his *Politics Among Nations: The Struggle for Power and Peace* the following analysis:

The history of modern political thought is the story of a contest between two schools that differ fundamentally in their conceptions of the nature of man, society, and politics. One believes that a rational and moral political order, derived from universally valid abstract principles, can be achieved here and now. It assumes the essential goodness and infinite malleability of human nature, and blames the failure of the social order to measure up to rational standards on the lack of knowledge and understanding, obsolescent social institutions, or the depravity of certain isolated individuals and groups. It trusts in education, reform, and the sporadic use of force to remedy these defects. The other school believes the world, imperfect as it is from the rational point of view, is the result of forces inherent in human nature. To improve the world one must work with those forces, not against them. This being inherently a world of opposing interests and of conflict among them, moral principles can never be fully realized but must at best be approximated through the ever temporary balancing of interests and the ever precarious settlement of conflicts (Morgenthau 1948, p. 3).

In international relations, the former school is commonly called political liberalism and the latter school is commonly called political realism. Political realists believe that nation-states are the primary actors in the realm of international relations, and that these nation-states ought to always act in accordance with their own self-interest. (Interest here is defined by political realists in terms of military power – either the offensive military power a nation-state has in comparison to its rivals or the defensive military power a nation-state has in comparison to its rivals. The former are called offensive realists, whereas the latter are called defensive realists). In particular, the political realist considers it the aim of every nation-state to accumulate military power in order to become a hegemon: “Given the difficulty of determining how much power is enough for today and tomorrow, great powers recognize that the best way to achieve hegemony now, thus eliminating any possibility of a challenge by another great power. Only a misguided state would pass up an opportunity to be the hegemon in the system because it thought it already had sufficient power to survive. But even if a great power does not have the wherewithal to achieve hegemony (and that is usually the case), it will still act offensively to amass as much power as it can, because states are almost always better off with more rather than less power” (Mearsheimer 2001, p. 35). This is precisely the system of international relations that Kant is describing in the following passage: “So long as a state has a neighboring one which it can hope to subdue, it strives to aggrandize itself by subjugating it. It thus strives fora universal monarchy – a state constitution in which all freedom would necessarily expire, and, together with it, virtue, taste and science (which follow upon freedom)” (6:34). In the *Religion*, therefore, Kant is describing the system of international relations the way an offensive realist would. This may be contrasted with Kant’s description of the system of international relations in *Perpetual Peace: A Philosophical Sketch*, which is considered a key text of political liberalism.
went away the Enlightenment thinkers imagined it. The advance of human civilization merely sublimated this brutality away from our neighbors, towards more distant enemies. Therefore, not only does entering into the state of civilization add more vices in the form of the vices of culture, but man retains all his old vices of the savagery of nature. Hence Kant dismisses philosophical chiliasm the same way he dismisses theological chiliasm, because one must be possible if the other is: “So philosophical chiliasm, which hopes for a state of perpetual peace based on a federation of nations united in a world-republic, is universally derided as sheer fantasy as much as theological chiliasm, which awaits for the completed moral improvement of the human race” (6:34).

If this is indeed the right approach to interpreting the woeful examples that the experience of human deeds parades before us, then it becomes evident that Kant’s argument for radical evil is not meant to be an empirical generalization. The problem with interpreting the argument for radical evil as an empirical generalization is that it makes Kant a terrible philosopher, because there is no way these examples can be used to prove that the propensity is universal. Allison explains this well when he observes: “But clearly, even if for the sake of argument, one accepts Kant's appeal to some rather selective anthropological evidence, the most that this evidence can show is that evil is widespread, not that there is a universal propensity to it” (Allison 1990, p. 154). But it is not necessary to interpret the argument for radical evil as Kant's attempt to prove the universality of radical evil through the use of this anthropological evidence. For Kant seems to be using this anthropological evidence to disprove the hypothesis that human beings are by nature good, and this would be a perfectly good use for them. Even though it is impossible to prove an empirical generalization in this way, no matter how many examples we choose to provide, the opposite is
not true – only one example is needed to disprove an empirical generalization. And Kant does precisely that. The examples of the ritual murders of Tofoa and the never-ending cruelty of northwestern America serve to disprove the empirical generalization that human beings are good when in the state of nature. And the examples of the different vices of culture and of realpolitik in the relations between nation-states serve to disprove the empirical generalization that human beings are good when in the civilized state. In other words, the woeful examples that the experience of human deeds parades before us are not meant to provide verification for the doctrine that human beings are by nature evil, but for the falsification of the doctrine that human beings are by nature good.

33 The positions described here are therefore reminiscent to those featured in the debate over the problem of induction in the philosophy of science. The problem of induction has many forms, but the problem we are interested in is the problem of justifying interests of the following form:

\[ a_1, a_2, \ldots, a_n \text{ are all } Fs \text{ that are also } G \]

Therefore, all \( Fs \) are \( G \)

The verificationist position is that we would need to verify that every object \( a_1 \) through \( a_n \) is an \( F \) that is also a \( G \). If we do this, and it is true that every object \( a_1 \) through \( a_n \) is an \( F \) that is also a \( G \), then we are justified in making the inference that all \( Fs \) are \( G \). But this method seems unpromising. It is not always possible to examine every \( F \) to verify that are \( G \). And it is also not always possible to know that the \( Fs \) we have examined are all the \( Fs \) in existence. For this reason, Popper proposed the falsificationist position. The assertion that all \( Fs \) are \( G \) is, according to Popper, a hypothesis. As such, it may be posited without the need for justification. If we examine every object \( a_1 \) through \( a_n \) and find that they are all \( Fs \) that are also \( G \), we have not verified the hypothesis that all \( Fs \) are \( G \). All we have done is avoided falsifying the hypothesis that all \( Fs \) are \( G \). As Popper puts it in *The Logic of Scientific Discovery*: “Scientific theories can never be ‘justified’, or verified. But in spite of this, a hypothesis \( A \) can under circumstances achieve more than a hypothesis \( B \) – perhaps because \( B \) is contradicted by certain results of observations, and therefore ‘falsified’ by them, whereas \( A \) is not falsified; or perhaps because a greater number of predictions can be derived with the help of \( A \) than with the help of \( B \). The best we can say of a hypothesis is that up to now it has been able to show its worth, and that it has been more successful than other hypotheses although, in principle, it can never be justified, verified, or even shown to be probable. This appraisal of the hypothesis relies solely upon deductive consequences (prediction) which may be drawn from the hypothesis. There is no need to even mention induction” (Popper 1959, p. 317).

Kant cites a multitude of woeful examples that the experience of human deeds parades before us as proof that human nature is evil. When interpreters of the *Religion* dismiss this proof because no amount of examples of human misdeeds, however numerous, is sufficient to prove the universality of radical evil, they have tacitly interpreted Kant’s argument as verificationist in nature. This is evident when they dismiss the anthropological evidence provided because it is not possible to examine every human being for moral wickedness. But perhaps we do not need to ascribe to Kant a verificationist approach. For then we would no longer assume that the anthropological evidence Kant has provided is supposed to justify his inference from the misdeeds of certain classes of human beings to the presence of evil in the species. Instead, we would interpret Kant as using this anthropological evidence to falsify competing hypotheses concerning the moral character of the human species. That is, we interpret Kant’s argument in the *Religion* in accordance with a falsificationist methodology and not a verificationist methodology.
Radical Evil as Innate Disposition

Radical evil is described as the propensity to evil throughout the *Religion*. It is therefore important to understand precisely what this term “propensity to evil” consists in. Kant provides us with two different explanations of what he means by a propensity in the section of the *Religion* where he introduces it. First, he describes a propensity as a species of predisposition: “Propensity is actually only the *predisposition* to desire an enjoyment which, when the subject has experienced it, arouses *inclination* to it” (6:29n). In his example, Kant cites the propensity for intoxicants that is present in savages, where “although many of them have no acquaintance at all with intoxication, and hence absolutely no desire for the things that produce it, let them try these things but once, and there is aroused in them an almost inextinguishable desire for them” (6:29n).34 But there is

34 Frierson discusses the structure of Kant’s moral psychology using the example of a savage with a propensity for intoxicants from the *Religion*. He notes that the causal model in a case of an inclination to strong drink would look like so:

```
Sensory cognition (sight or smell of strong drink) → Feeling/Desire (desire to consume drink)
↑ Inclination (for strong drink)
```

But Kant considers inclinations to be acquired. Inclinations are acquired through past experiences, which subsequently develop into a habitual desire:

```
Sensory cognition (sight or smell of strong drink) → Feeling/Desire (desire to consume drink)
Past experience with strong drink ↑ Inclination (for strong drink)
```

This causal connection between the subject’s past experience with strong drink and the inclination for strong drink needs to be explained. In other words, a reason why some (but not all) experience develop into habitual desires. And this introduces the concept of a *propensity* into the causal model:
also the description of propensity that makes it into a ground of inclinations: “By *propensity* (propensio) I understand the subjective ground of the possibility of an inclination (habitual desire, *concupiscentia*), insofar as this possibility is contingent for humanity in general” (6:29).\(^{35}\) A propensity here described is not a predisposition. For this passage tells us that predispositions are innate and can only be represented as innate. But a propensity is described here as a ground “that can indeed be innate yet *may* be represented as not being such: it can rather be thought of (if it is good) as *acquired*, or (if evil) as *brought* by the human being *upon* himself” (6:29). Thus, here propensities and predispositions are being described as distinct species of innate dispositions, and their difference consist in the fact that the former can be represented as freely acquired, whereas the other cannot be so represented.

But what are dispositions? In the contemporary discussion of properties, a property can either be categorical or dispositional. The standard way of making the distinction has been to say that dispositional properties contain a conditional, whereas categorical properties do not. For example, a paradigmatic categorical property would be the shape of an object. An object’s surface

\[
\begin{array}{ccc}
\text{Sensory cognition} & \rightarrow & \text{Feeling/Desire} \\
(\text{sight or smell of} & & (\text{desire to} \\
\text{strong drink}) & \uparrow & \text{consume drink}) \\
\text{Past experience with} & \rightarrow & \text{Inclination} \\
\text{strong drink} & & (\text{for strong} \\
& & \text{drink}) \\
\uparrow & & \text{Propensity (for} \\
& & \text{strong drink)}
\end{array}
\]

In this model, our desires are grounded in inclinations, and our inclinations are grounded in turn in propensities and predispositions (Frierson 2005, pp. 21-22). The location of the propensity in the causal model as the subjective ground to another subjective ground (the inclination) demonstrates what Kant had in mind when he makes the following statement defining propensity: “*Propensity* is actually only the *predisposition* to desire an enjoyment which, when the subject has experienced it, arouses *inclination* to it” (6:29n).

\(^{35}\) It should be noted that Kant provides yet a *third* definition of propensity in the *Anthropology*, where he defines it as: “The subjective *possibility* of the emergence of a certain desire, which *precedes* the representation of its object, is *propensity* (propensio)” (7:265). This definition is similar to, but not the same as, the definition of propensity as the subjective ground of the possibility of an inclination (6:29). This is because an inclination is a desire that is *habitual*. The *Anthropology* definition of propensity therefore encompasses a wider range of cases (all desires) than the *Religion* definition of propensity (only those desires that are habitual).
is square when all of its sides are of the same length. By contrast, a paradigmatic dispositional property would be fragility. For when we say that an object is fragile, we mean that if it was subjected to a sudden force, it would shatter. And here we have a condition (subjection to a sudden force) and the result that occurs when this condition is met (the object shattering into pieces). There are two takeaways from this that are relevant here. First, insofar as dispositional properties contain a conditional, it follows that the disposition may never manifest itself because this...

36 The existence of dispositional properties is by no means a settled matter. In the context of contemporary Analytic philosophy, the debate on dispositional properties can be traced to the logical empiricists, who objected to dispositional properties because of their semantic indeterminacy. For the logical empiricists, the issue with dispositional properties is this: ‘Imagine hesitating to attribute the dispositional predicate ‘is water-soluble’ to a given object. If the object undergoes the appropriate test, in this case, if it is put into water, all is well, because in the circumstances of the test, the application conditions of the predicate are not conditional: while it undergoes the test, the object is water-soluble if it dissolves and not water-soluble if it does not dissolve. The problem arises in circumstances in which the object under consideration does not undergo any appropriate test. What determines in this case whether the object is water-soluble?” (Kistler-Gnassounou 2007, p. 24). The logical empiricists give the following answer in response to this: there is nothing that determines whether the object is water-soluble under those circumstances. This means that it is up in the air whether the object is water-soluble or not, i.e. the property “is water-soluble” is an indeterminate property.

The idea that an object can have indeterminate properties is obviously problematic. The positions that subsequent philosophers have adopted to avoid adopting the logical empiricist position can be divided into roughly two camps. There are philosophers who, like Quine, deny that there are dispositional properties entirely: “Quine’s thesis is that the attribution of a dispositional predicate to an object reflects our incomplete knowledge of that object: when we call the sugar ‘soluble’, we in fact attribute a categorical property, though one that we know only incompletely. The real properties making true attributions of dispositions are categorical properties of ‘microstructure’. By using a dispositional predicate, says Quine, ‘we can refer to a hypothetical state or mechanism that we do not yet understand’. From his point of view, to attribute a disposition to an object is just a provisional way to characterize it faute de mieux, awaiting a more direct way. There are really no dispositional properties, the dispositional being only an incomplete and provisional way to characterize microstructural categorical properties” (Kistler-Gnassounou 2007, p. 27). For Quine, there is something that determines whether the object is water-soluble under the circumstances described, and that is the categorical properties of microstructure. If the object has a certain microstructure, then the object is water-soluble, and if it does not have this microstructure, then it is not water-soluble.

Then there are philosophers who follow Goodman in accepting the existence of dispositional properties. As such, they attempt to respond to Carnap’s objection by understanding dispositional properties in terms of counterfactual conditionals: “... This is the research tradition initiated by Goodman: his starting point is the observation that dispositions ‘strike us by comparison [with observable properties] as rather ethereal. Ans so we are moved to inquire whether we can bring them down to earth; whether, that is, we can explain disposition-terms without any reference to occult powers.’ According to Goodman’s hypothesis, which constitutes until today the framework of this debate, the meaning of dispositional predicates can be analyzed in categorical terms with the help of counterfactual conditionals” (Kistler-Gnassounou 2007, p. 26). For Goodman, there is something that determines whether the object is water-soluble under the circumstances described, and that is the counterfactual properties the object has. If the object would dissolve if it is immersed in water, then it has the dispositional property of being water-soluble, even if it never actually undergoes the test of being immersed in water. And if the object wouldn’t dissolve if it is immersed in water, then it lacks the dispositional property of being water-soluble, even if it never actually undergoes the test of being immersed in water.

For a more in-depth discussion of these two positions on the existence of dispositional properties, see Dispositions and Causal Powers (2007) by Max Kistler and Bruno Gnassounou.
condition is never met. As an example, consider a glass that is fragile. This means that, if the glass is subjected to a sudden force, then it would shatter. But if this glass is never subjected to a sudden force, then it would never have the occasion to shatter, unless we posit further conditions that would also result in its shattering (for example, a high-pitched sound). Second, an object has the dispositional property whether or not it actually manifests. To return to our example of a fragile glass, we say that the glass is fragile because, if the glass is subjected to a sudden force, it would shatter. But a glass that is never subjected to a sudden force remains fragile, even though it never has and never will have the occasion to shatter. The glass retains its dispositional property of fragility on the basis of the following counterfactual’s being true: Even though the glass was never subjected to a sudden force, had it been subjected to a sudden force, it would have shattered into pieces.

There is evidence that Kant thinks about dispositions along these same lines. In the following passage from the Metaphysik Herder, Kant writes: “If a substance suffers, then it must contain in itself by its own power the ground of the inherence of the accident, because otherwise the accident would not inhere in it” (28:51-52). From this, it follows that it is not enough to have a determining ground that allows us to posit a predicate in the subject. There must also be a corresponding ground in the subject that allows this predicate to inhere in the subject. Kant’s reasoning on the matter is that, if a determining ground was sufficient by itself to produce a determination in a subject, then it would be possible for a determining ground of sufficient power (God) to make a wooden post think (28:52). But this is plainly absurd, because even if God could produce a thought in the wooden post all by himself, it would be God, and not the wooden post, that has the thought: “Thus for the inherence of an accident in A its own power is required, and a
merely external, not even a divine power, does not suffice” (28:52). To put this in the terms of our discussion in the previous paragraph, consider a glass that has the dispositional property of being fragile, and a metal block that does not. Now we say something is fragile if it is disposed to shatter into pieces when subjected to a sudden force. Therefore, because glass is fragile and metal is not, when subjected to such a force, the glass shatters but the metal block does not. Now God, by his power, could cause the metal block to shatter into pieces, but this does not make the metal block *fragile*. And the reason why it does not is because it lacks the relevant dispositional property. The metal block is not *disposed* to shatter when struck by a sudden force, it is not *characteristic* of metal blocks to behave in this way (even though they conceivably could so shatter due to divine intervention).

If we are right in our reading of Kant, then he intends for radical evil to be a *dispositional property* in the same sense we do when we use the term. Now to repeat what a propensity is: “By *propensity* (*propensio*) I understand the subjective ground of the possibility of an *inclination* (habitual desire, *concupiscentia*), insofar as this possibility is contingent for humanity in general” (6:29). The propensity *to evil* would therefore have to be the subjective ground of the possibility of an inclination *to evil*, insofar as this possibility is contingent for humanity in general. And as a dispositional property, the propensity to evil would contain a condition for the manifestation of

37 In his article “The Critique of Metaphysics: Kant and Traditional Ontology”, Ameriks discusses this argument as the Restraint Argument. He writes: “Like Baumgarten, Kant wants to argue from the start that action is always a mixture of spontaneity and reaction, and that in any real action there are always several concurring causes. For example, when we listen with attention, outer things are a true ground of the experience; but, in attending, we are also playing a role, so we are active and passive at once. In particular, Kant stresses that, even for God to put a thought into us, there must be a ground within us, a capacity to receive and have the thought; otherwise, there would be no point in saying that it is we rather than God who have the thought” (Ameriks 2003, p. 125). Ameriks calls this argument the Restraint Argument because it prevents us from ascribing *all* activity and reality to God (Ameriks 2003, p. 125). This application of the Restraint Argument to the actions of God must not, however, cause us to overlook Kant’s point: In *any* real action, there are always several concurring causes. Therefore, for any external cause (even God) to bring about a state in us, there must be a ground within us that has the capacity to receive that state. This ground would, in contemporary parlance, be called a *disposition*. 
this inclination.\textsuperscript{38} For example: a human being has the propensity to evil because he would succumb to evil if he were subjected to a great enough temptation. Here we have a condition (subjection to temptation) and a result that occurs when the condition is met (succumbing to evil). The two important takeaways from before are still relevant here. First, insofar as dispositional properties contain a conditional, it follows that the disposition may never manifest because the condition is not met. This means that, even though a human being might have the dispositional property of moral frailty, he might never succumb to evil because he has never encountered temptation (or never encountered great-enough temptation). The human being that lives his life without ever having experienced temptation would never have the occasion to succumb to the temptation. Second, an object retains its dispositional properties whether or not they manifest. This means that the human being that has never had the occasion to succumb to temptation because of the temptation-less life he’s led still has the dispositional property of moral frailty. He retains the dispositional property of moral frailty based on the truth of the counterfactual that, if he is subjected to a great enough temptation, then he would falter and he would succumb to the temptation.\textsuperscript{39} For

\textsuperscript{38} As Mariña notes, the human being’s choice of a fundamental maxim determines his orientation to the world. For the individual that chooses to adopt a fundamental maxim that is evil, he develops a propensity to subordinate the moral incentives to other incentives that are nonmoral. It would be inaccurate to describe this as depravity, however, if what we mean by it is a conscious decision to reverse the ethical order of incentives (6:30). The influence of the propensity is far more subtle: “... Because the individual has already made a grounding choice – too often unexamined – about what she takes to be most important, she will simply find the objects of her daily concern mattering to her in one way or another. The propensity is so deeply ensconced in the soul that it preexists experience and hence actual incentives. In fact, a person may not have certain inclinations only because they have not yet awoken. As such, it may be quite accidental that a person is not subject to certain vices, for she simply have not encountered the conditions that would have brought them out. The subjective ground for them is, however, still there. And because at the level of her inner phenomenal life she simply finds things she encounters mattering to her in a certain way, this has enormous effects on her imaginative capacities, on what characteristics of a situation she will identify as salient, on what, for her, counts, and hence, on which moral demand she considers to be at play. In other words, the propensity conditions how an individual understands and interprets her situation and its demands” (Mariña 2017, pp. 196-197).

\textsuperscript{39} The argument that Kant makes in 6:39 appears to have more in common with the rhetorical enthymeme than the formal syllogism. Traditionally, the enthymeme has been defined in the following manner: “The Enthymeme must consist of few propositions, fewer often than those which make up the normal syllogism. For if any of these propositions is a familiar fact, there is no need to even mention it; the hearer adds it himself” (Rhetoric 1357a16-18). For an example, consider the formal syllogism:
this reason, we should take note of the fact that Kant calls the propensity to evil a potentiality for sin: "... When we enquire into the origin of evil, at the beginning we still do not take into account the propensity to it (as peccatum in potentia) but only consider the actual evil of given actions according to the evil's inner possibility, and according to all that must conspire within the power of choice for such actions to be performed" (6:40-41).

Let us now try to reconstruct Kant's reasoning that leads him to assert that radical evil is universal among human beings. There are three points to consider here. First, evil is a possibility for the entire race. This is where Kant's anthropological research enters the picture. For this anthropological research is not only useful in disproving the competing hypotheses on the moral bent of the human species, but the same research also serves to show that this evil cannot be confined to any sub-section of the human species: every human being is capable of evil-doing. Second, this possibility can only be explained through the presence of a dispositional property

All men are mortal
Socrates is a man

Therefore, Socrates is mortal

From this formal syllogism, an enthymeme may be constructed by leaving either the major premise or the minor premise unstated. If it is the major premise that is left unstated, then the enthymeme we have constructed will state: Socrates is a man; therefore he is mortal. If it is the minor premise that is left unstated, then the enthymeme we have constructed will state: All men are mortal; therefore Socrates is mortal.

Kant writes in 6:39 the following: "A member of the English Parliament exclaimed in the heat of debate: "Every man has his price, for which he sells himself". If this is true (and everyone can decide by himself), if nowhere is a virtue which no level of temptation can overthrow, if whether the good or evil spirit wins us over only depends on which bids the most and affords the promptest pay-off, then, what the Apostle says might indeed hold true of human beings universally, "There is no distinction here, they are all under sin – there is none righteous (in the spirit of the law), no, not one." (6:38-39). This would give us the following:

If "Every man has his price, for which he sells himself", then there is a universal propensity to evil

This would form the major premise of the syllogism. The minor premise and conclusion are both left out by Kant here (to be supplied by the reader). This makes the argument in 6:39 an enthymeme (rhetorical syllogism) rather than a formal syllogism. But it is evident that the formal syllogism Kant had in mind can be reconstructed so:

If "Every man has his price, for which he sells himself", then there is a universal propensity to evil
It is true that "Every man has his price, for which he sells himself"

Therefore, there is a universal propensity to evil
belonging to the entire human species. By describing radical evil in terms of a propensity, Kant makes it a disposition that is innate to all human beings. Insofar as the propensity to evil is the disposition to succumb to evil when subjected to great enough temptation, it is not necessary for Kant to prove that every human being has succumbed to evil, because it is possible that there are human beings for whom the condition has not been met. What is important here is that the human being in question has not succumbed to evil as a matter of fact, not as a matter of necessity.  

Third, given that every human being has this innate disposition, it must be concluded that human nature is evil, the original predisposition to good notwithstanding. The reason for this is because ethical rigorism has ruled out the possibility that human nature might be morally neutral and the possibility that human nature might be partly good and partly evil. And while experience inclines us towards ethical latitudinarianism, if we had to describe human nature as either good or evil, we would have to describe human nature as evil, because even the smallest imperfection mars the beauty that is moral perfection: For whosoever keeps the whole law, yet offend at one point, he is guilty of all (James 2:10).

Kant raises the possibility of sages with finite holy wills in the Metaphysics of Morals, for example, who might be examples of human beings who never succumb to evil (6:383). As representatives of human morality in its highest stage, the sage would be an example of a human being who has restored his original predisposition to good (6:46). And succumbing to temptation would be impossible for the sage on account of this purity, that is, being free from the influence of any incentive other than that of duty (6:383).

In this reading, the argument for radical evil bears an interesting resemblance to transcendental arguments as interpreted by Karl Ameriks. Ameriks interprets transcendental arguments as regressive arguments. In reading the argument of the Analytic as a regressive argument, Ameriks ascribes to it the following structure: “... Empirical knowledge (‘experience’) is possible only if the ‘original synthetic unity of apperception’ applies to it, which is possible only if pure concepts have validity, and this in turn requires that transcendental idealism be true. Such an argument would be transcendental in that it too would explain that a body of knowledge is possible only if there are representations (pure concepts) of a certain kind” (Ameriks 2003, p. 54). This is to be contrasted with the progressive approach: “The major departure of this interpretation is that it takes the Critique to accept empirical knowledge as a premise to be regressively explained rather than as a conclusion to be established. Peter Strawson, Jonathan Bennett, and Robert Paul Wolff have insisted at length that such an argument is undesirable, uninteresting and not representative of Kant’s best intentions. They all represent the transcendental deduction as basically aiming to establish objectivity, i.e. to prove that there is an external and at least partially lawful world, a set of items distinct from one’s awareness, and to do this from the minimal premise that one is self-conscious” (Ameriks 2003, p. 55). Likewise here, we begin with the idea that moral evil in human beings is a premise to be regressively explained rather than a conclusion to be established. From here, we posit the existence of radical evil as the condition for the possibility of the moral evil in human beings. This is different in comparison to other interpretations of the argument for radical
Frailty and Depravity

A problem arises at this point. Kant refers to the evil in human nature exclusively as radical evil. This obscures the fact that radical evil can be understood as either a fundamental maxim or as a propensity. Radical evil, as a fundamental maxim that is evil, is the intelligible ground of all our (empirical) maxims. It determines the form that our maxims take. For a maxim is good or evil depending on whether we subordinate the moral incentive or the nonmoral incentive that we have incorporated into our maxims. If the fundamental maxim of the human being is good, then he will take the moral law as the sole sufficient determining ground of his maxims, and the nonmoral incentive will invariably be subordinated to the moral incentive in his maxims. But when the fundamental maxim of the human being is evil, then the moral law is merely one among the many grounds that can determine his maxims, and this makes possible the subordination of the moral incentive to the nonmoral incentive in his maxims. But as the propensity to evil, radical evil is the phenomenal ground of all our wayward inclinations. It is (partially) the source of the matter of our maxims, insofar as our inclinations determine what can serve as incentives for us. These incentives (moral and nonmoral) are then incorporated into our maxims, and our maxims are determined to be good or evil depending on which incentive is given priority in the maxim. Therefore, absent the propensity to evil, a real opposition between the moral incentive and the nonmoral incentive would be impossible, and the human being’s power of choice would always act on the moral incentive it has incorporated into the maxim. Now there would be no problem if either conception of radical evil can be reduced to the other, because then Kant would just be speaking of one thing in two ways, and not of two distinct things. As we will demonstrate in this section, however, neither evil that argue progressively from a condition that is present in every human being, e.g. the embodied nature of human beings (Michelson) or the social nature of human beings (Wood), and prove that radical evil is present in every human being on account of their having this condition. It is after they have proved that radical evil exists that they accept the moral evil in human beings.
conception of radical evil is reducible to the other. For the evil of depravity is grounded in a fundamental maxim that is evil. Therefore, if we take radical evil to be a propensity only, depravity becomes incomprehensible as an evil. Yet the opposite course of action is also closed to us. If we take radical evil to be a fundamental maxim only, then frailty becomes incomprehensible as an evil. For the evil of frailty does not consist in any maxim, but in the real opposition of inclinations that are able to prevent a maxim from being efficacious. Radical evil seems, therefore, to be a patchwork of disparate ideas, connected to each other by nothing more than the empirical deeds that they are supposed to ground.42

42 For Palmquist, radical evil is to be sharply distinguished from the propensity to evil. He writes: “Although the term ‘radical evil’ appears in the title of Book One, Kant mentions it only a few times in the main text itself. Nevertheless, it should not be regarded as a mere synonym for the evil propensity. Rather, it is Kant’s technical name for the mysterious noumenal origin of our propensity to evil. Naming it in this way does not actually explain it, so much as indicate the extent of our ignorance of its origin” (Palmquist 2000, p. 157n). Neither is radical evil identical with the fundamental maxim, because according to Palmquist, the propensity to evil is the cause of the choice of an evil fundamental maxim, and not the result of this choice (Palmquist 2000, p. 158). Palmquist’s account of radical evil can be broken down into three steps.

In the first step, the good predisposition provides a person with the potential to do good (Palmquist 2000, p. 151). This is represented in the following manner:

\[
\text{good predisposition} \rightarrow \text{original innocence} \rightarrow \text{potential to do good}
\]

In the second step, radical evil corrupts the potential to do good into the propensity to evil (Palmquist 2000, p. 158). Palmquist does not explain how this corruption takes place, but is content in following Kant in saying that this corruption is contingent (and therefore free) and universal to every human being. Hence:

\[
\text{potential to do good} \rightarrow \text{radical evil} \rightarrow \text{propensity to evil}
\]

In the third step, Palmquist describes what happens as: “The choice of an evil maxim as supreme realizes the human propensity to evil by producing an evil heart” (Palmquist 2000, p. 158). And this is represented as:

\[
\text{propensity to evil} \rightarrow \text{act of choosing an evil supreme maxim} \rightarrow \text{an evil heart}
\]

This means that Palmquist’s reading does not have a difficulty with the tension between frailty and depravity, because he does not trace frailty to the propensity to evil and depravity to the evil fundamental maxim. The three gradations arise at Step 2 of his process, because they are, according to Palmquist, the corruption of the good potentialities of the human being (Palmquist 2000, pp. 152-153). The predisposition to animality is corrupted by radical evil into frailty, or a self-love rooted in sensibility. The predisposition to humanity is corrupted by radical evil into impurity, or a self-love rooted in understanding. And the predisposition to personality is corrupted by radical evil into wickedness, or a self-love rooted in judgment. It seems, however, that this interpretation of the origins of the three
Let us consider why radical evil, considered as a propensity, is unable to ground empirical deeds of depravity. Fundamentally, depravity has to do with our maxims, as can be seen in Kant’s account of depravity: “… The depravity (*vitiositas, pravitas*) or, if one prefers, the *corruption* (*corruption*) of the power of choice to maxims that subordinate the incentives of the moral law to others (not moral ones)” (6:30). In depravity, the *Willkür* freely chooses to subordinate the moral incentive to nonmoral incentives, and this subordination is what makes the *Willkür’s* choice an evil one. Per the incorporation thesis, we know that the *Willkür* makes this choice by incorporating the relevant incentive into its maxim: “… Freedom of the power of choice has the characteristic, entirely peculiar to it, that it cannot be determined to action through any incentive *except so far as the human being has incorporated it into his maxim* (has made it into a universal rule for himself, according to which he wills to conduct himself)” (6:24). It is therefore safe to say that the incentives play no role in Kant’s account of what it means to act depravedly. The problem is that for Kant a propensity is a subjective ground for the possibility of an inclination or a habitual desire (6:29). And these inclinations and habitual desires become *incentives* for us to act, albeit not moral ones. It follows then that the propensity to evil grounds our nonmoral incentives to act in a manner contrary to the moral law. But according to the incorporation thesis, the incentives do not determine the power of choice to action: it is *only* determined to action through the incorporation of one of these incentives into the maxim. The propensity therefore arrives at the scene too early, before any wrongdoing has even occurred. Depravity places the moral responsibility of wrongdoing squarely on the *Willkür*, and we do not choose whether or not we *possess* one incentive.

Gradations of evil is inconsistent with Kant’s declaration in 6:28, where he argues that the predisposition to personality cannot be corrupted in the way Palmquist envisions, otherwise the moral law would be unable to serve as an incentive for the human being: “… A power of choice so constituted is a good character, and this character, as in general every character of the free power of choice, is something that can only be acquired; yet, for its possibility there must be present in our nature a predisposition onto which nothing evil can be grafted” (6:27 – see also 6:28).
or the other, so much as we choose how to subordinate the incentives that we do possess, and so there seems to be a fundamental mismatch between the evil of depravity and its ground in radical evil.

Let us consider why radical evil, considered as a fundamental maxim, is unable to ground the empirical deeds committed out of frailty. The starting point of our discussion here is a passage that comes later in the Religion, in which Kant writes: “The original good is holiness of maxims in the compliance to ones’ duty, whereby a human being, who incorporates this purity into his maxims, though on this account still not holy as such (for between maxim and deed there is still a wide gap), is nonetheless upon the road of endless progress towards holiness” (6:46-47). It is surprising to see Kant say that there is a wide gap between the maxim and the deed. For consider the implications of this assertion: According to the incorporation thesis, an incentive must be incorporated into a maxim in order for the will to be determined to action. But once this incentive has been incorporated into the maxim, how is it possible that the will remains undetermined with respect to action? If the transition from the maxim to the deed is liable to fail, then there must be a reason for this. That is, we need to posit the existence of an obstacle that needs to be overcome before the maxim can be translated into action. But there isn’t any obstacle that can impede our maxims in this way except for our feelings and inclinations, because no other counterweight exists within us that can stand up to the moral law in this manner, as Kant observes in the following passage from the Groundwork: “The human being feels within himself a powerful counterweight to all the commands of duty, which reason represents to him as so deserving of the highest respect – the counterweight of his needs and inclinations, the entire satisfaction of which he sums up under the name happiness” (4:405). But the incorporation thesis has already ruled out these feelings and inclinations, because to have reached this point in our reasoning process means that the incentives
of sensibility have already been rejected in favor of incorporating the incentive of the moral law into our maxim to form the good maxim. The feelings and inclinations therefore seem to play a double role, not only as the incentives which, when incorporated into a maxim, allow for an act of depravity, but also as an additional obstacle that can prevent a good maxim from becoming a good deed.

It would be easy to dismiss this separation between the maxim and the deed if the passage above was just an isolated incident. But this way out is not available to us, because there are crucial aspects of Kant’s account of radical evil that assume just such a separation of the maxim from the deed. The discussion of frailty, which is represented as a grade of the propensity to evil, proceeds in the following way: “... The frailty (fragilitas) of human nature is expressed even in the complaint of an Apostle: “What I would, that I do not!” i.e. I incorporate the good (the law) in to the maxim of my power of choice; but this good, which is an irresistible incentive objectively or ideally (in thesi), is subjectively (in hypothesi) the weaker (in comparison with inclination) whenever the maxim is to be followed” (6:29). This passage is important for the argument we are making in this section, since it shows that, according to Kant’s analysis of frailty, there is only one maxim present in a given case of frailty, and since the moral law has been prioritized in this maxim as its incentive, the only maxim present in a given case of frailty is a good maxim. But even though the only maxim present in a given case of frailty is a good maxim, it is somehow possible for the moral incentive to be resisted, and the good deed that is supposed to follow from the good maxim never materializes, from which results the evil of frailty. The account of frailty makes it clear that an action prescribed by our maxim need never come to pass in deed, and insofar as frailty is represented as a weakness for desires and inclinations that are contrary to the moral law, it also
confirms our earlier speculation that the inclinations are the obstacles that prevent our chosen maxim from becoming a deed.

This account of frailty is not easily reconciled with the idea that evil lies in our maxims. Frailty is not an evil that can be represented in terms of a choice of an evil maxim. Here we are not confronted with a choice between the moral incentive and nonmoral incentives. In the case of frailty, this choice needs to have already happened, in which we have chosen to subordinate the nonmoral incentive in our maxim to form a good maxim.\(^{43}\) For it is only after we have made this choice to subordinate the nonmoral incentive, and constructed a good maxim, that it becomes possible for the good maxim to fail to become a good deed. But this failure cannot be explained by positing an opposing maxim. This is because a maxim is a general principle for our actions, by reference to which we will our conduct (6:24). Therefore, if our power of choice is determined to \(\Phi\), it is because it falls under a universal rule we have constructed for ourselves, according to which we have willed to conduct ourselves. And if our power of choice is determined to \(\not\Phi\), it is because it falls under a different universal rule that we have constructed for ourselves, according to which we have willed to conduct ourselves. This means that if the reason a good maxim failed to become a good deed consists in the interference of an opposing maxim, then there would have to be two universal rules in us, and that we have willed to conduct ourselves in two opposing ways. But how then would we be able to break the stalemate between the opposing maxims that direct

\(^{43}\) Baron makes a similar observation regarding frailty in the following passage: “Moral weakness involves resolving to do something which one believes is morally required, and failing to act accordingly (without any change of heart as to what one should do). But how is this possible, on the model of human agency that Allison attributes to Kant? It cannot be that the agent is overcome by desire, since according to the Incorporation Thesis, no desire can move the agent unless she incorporates it into her maxim. The Incorporation Thesis would seem to require a denial of the possibility of weakness” (Baron 2008, p. 433). For the entire point of the Incorporation Thesis is to deny the possibility of the human being’s power of choice being moved to action (and presumably inaction also) by an incentive (6:24). The human being must have input on what determines him to action by incorporating the incentive into his maxim (6:24). Whereas moral weakness, as Baron understands it, represents the human being as being overcome by desire, and this seems antithetical to the point Kant is trying to make in the Incorporation Thesis.
us to $\phi$ and to not $\phi$? It would require an appeal to a maxim that is higher up in the human being’s hierarchy of maxims to break the stalemate. But every maxim has its ground in a higher maxim. Therefore, if we have a maxim that directs us to $\phi$ and a maxim that directs us to not $\phi$ as our maxims, then both of them must be grounded in a higher maxim, and since they cannot both belong to the same series, they would have to each be grounded in a different first subjective ground for the adoption of maxims, which is absurd, because different fundamental maxims cannot coexist in the same subject. And if the good maxim is not opposed by an evil maxim, but is instead opposed by inclinations that interfere with the designs of our good maxims, then it would be useless to posit the formation of another good maxim tasked with overcoming these opposing inclinations. For this maxim would need to become a deed in order to be efficacious in overcoming the opposing inclinations, but insofar as there is a wide gap between the maxim and the deed, this maxim can also be prevented from becoming a deed by the interference of our desires and inclinations, and the regress would continue unto infinity. And since this regress is never resolved, the gap between maxim and deed remains unbridged, and no maxim belonging to the series will ever become a deed.

Allison is not blind to the problems that frailty creates for the account of evil associated with the incorporation thesis. This is why he addresses frailty specifically in the following passage: “… The so-called lack of sufficient strength to follow moral principles when they conflict with the claims of inclination reflects the lack of full commitment to these principles in the first place. Thus, self-deception enters the picture at the very beginning, depicting what is in reality a free evaluation on one’s part as a “weakness” for which one is not responsible” (Allison 1990, p. 159). But this proposal is not a solution to the problem so much as an attempt to make the problem go away, because the distinction between frailty and depravity no longer exists after Allison is done with it,
insofar as all that is left is a depravity that involves self-deception (or frailty) and a depravity that does not (or depravity per simplicitur). In other words, this proposal reinterprets the concept of frailty, in which Allison’s “basic hermeneutical assumption is that Kant’s entire account of evil, including the brief discussion of frailty, must be understood in light of the governing principle that “Man himself must make or have made himself into whatever, in a moral sense, whether good or evil, he is or is to become” (Rel 6:44; 40)” (Allison 1996, p. 120). This is however baldly contradicted by what Kant himself actually says about frailty, which is that the “irresistible incentive objectively or ideally (in thesi), is subjectively (in hypothesi) the weaker (in comparison with inclination) whenever the maxim is to be followed” (6:29). This distinction is not preserved when frailty is turned into depravity, which “reverses the ethical order as regards the incentives of a free power of choice” (6:30). The point is that frailty involves the formation of a good maxim, as Kant says “I incorporate the good (the law) into the maxim of my power of choice” (6:29), and this is incompatible with its being depravity, insofar as depravity involves an evil maxim, or “the propensity of the power of choice to maxims that subordinate the incentives of the moral law to others (not moral ones)” (6:30), which is an altogether different matter than self-deceptively treating the temptations we face as directly determining our power of choice to action, contra the incorporation thesis.

**Conclusion**

There is a formal proof for the universality of radical evil, and it is exactly where it purports to be. The multitude of woeful examples that the experience of human deeds parades before us proves that moral evil is possible for every human being, even the best among us. The propensity to evil has to be posited to ground this possibility, as a condition for the possibility of moral evil in every human being. But this development is bittersweet, because the conception of radical evil
as propensity is distinct from the conception of radical evil as fundamental maxim. Now Kant considers the empirical character of a human being to be grounded on his intelligible character: “… But the very same subject, being on the other side conscious of himself as a thing in itself, also views his existence insofar as it does not stand under conditions of time and himself as determinable only through laws that he gives himself by reason; and in this existence of his nothing is, for him, antecedent to the determination of his will, but every action – and in general every determination of his existence changing conformably with inner sense, even the whole sequence of his existence as a sensible being – is to be regarded in the consciousness of his intelligible existence as nothing but the consequence and never as the determining ground of his causality as a noumenon” (5:97-98). The empirical character of a human being is multifaceted. It includes his inclinations and desires, but also the beliefs he holds and the hopes he wishes for. The propensity to evil is only one facet of a human being’s character. But the choice of the fundamental maxim constitutes the intelligible character of a human being. Therefore it must be concluded: an evil fundamental maxim is the condition for the possibility of the propensity to evil, albeit as nothing more than a trivial metaphysical corollary of the proposition that the world of understanding contains the ground of the world of sense and so too of its laws (4:453). And if we are to move beyond this purely formal understanding of the grounding of the propensity to evil, we must give flesh to the bare-bones concept of a fundamental maxim through the moral-psychological terms of Kant’s practical philosophy.
CHAPTER 3. THE NOUMENAL ORIGINS OF RADICAL EVIL

Be ye therefore perfect, even as your Father which is in heaven is perfect.

Matthew 5:48

Introduction

How are we to prove that radical evil is really possible? The most controversial aspect of Critical philosophy is its transcendental idealism, in which Kant separates the appearances from the things themselves. Everything that is intuited in space and time is mere appearance, and thereby relegated to the sensible world. And everything that exists independently of us and our sensibility is a thing in itself, and thereby relegated to the intelligible world. But the concept of radical evil stands astride these two worlds, with one foot in the sensible world and the other in the intelligible world. This is because the concept of radical evil, considered as a fundamental maxim, is distinct from the concept of radical evil, considered as a propensity. As a fundamental maxim, radical evil is described by Kant as the subjective ground of the adoption of maxims (6:25). It is the intelligible ground of all our maxims, hence it is assigned to a place in the intelligible world. As a propensity, however, radical evil is described by Kant as the subjective ground of the possibility of an inclination (6:29). It is the sensible ground of all our wayward inclinations, and hence it is assigned to a place in the sensible world. The real possibility of radical evil consists in being able to bring together these two disparate concepts that Kant has subsumed under the heading “radical evil”, first by explaining how an evil fundamental maxim is even possible without reference to the sensible world, and second by explaining how the fundamental maxim is the intelligible ground of the natural propensity to evil.
It is by understanding what the noumenal choice of the radical evil consists in that its real possibility may be ascertained. To do so, the following question must be asked: How is the choice of a fundamental maxim that is evil even possible? For moral evil involves the incorporation of a sensible incentive, which is unavailable to us, as atemporal first causes making the choice of our fundamental maxims. Indeed, since the only incentive available to us is the moral one, we should all have fundamental maxims that are good, yet we do not. The only way out of this conundrum is to say that we incorporate the only incentive that is available to us, the moral incentive. The human being cannot escape the moral law that is engraved upon his heart. His choice consists in the attitude he adopts with respect to the moral law, which he can either embrace wholeheartedly, or accept begrudgingly. The wholehearted embrace of the moral law represents a moral strength of will that would never waver. While the mere acceptance of the moral law testifies to a moral strength of will that is imperfect, which will not remain steadfast before the obstacles it will face. And the human being cannot avoid facing these obstacles, which are placed before him because they belong with necessity to the possibility of his being.

**First**, the human being is always embodied. And because he is embodied, the human being will experience the sensuous impulses that come with having a body. The human being desires happiness: “… There is, however, one end that can be presupposed as actual in the case of all rational beings (insofar as imperatives apply to them, namely as dependent beings), and therefore one purpose that they not merely could have but that we can safely presuppose they all actually do have by a natural necessity, and that purpose is happiness” (4:415). The moral weakness of a human being in the face of obstacles of this sensuous nature is frailty. **Second**, the human being can never be without his humanity. And this predisposition to humanity in him gives rise to a conception of happiness that is not merely sensuous, which animals also have, but also
comparative: “The predispositions to humanity can be brought together under the general title of a self-love which is physical and yet involves comparison (for which reason is required); that is, only in comparison with others does one judge oneself happy or unhappy” (6:27). The problem with self-love is not that the human being desires happiness for himself, but that he has a tendency to put this desire for happiness above the dignity of his fellow man. Self-love has the tendency to turn into self-conceit. And the moral weakness of the human being that allows self-love to be corrupted into self-conceit is depravity. Frailty and depravity are not disparate concepts that Kant has subsumed under the heading “radical evil”. Radical evil is the moral weakness of will that is the condition for the possibility of frailty and depravity. The human being freely determines for himself his relationship with the moral law, and in so doing, inevitably falls to frailty and depravity.

**Noumenal Choice**

Every moral evil has to be a deed, and radical evil is no exception. But there is a considerable difference between radical evil and the evil deeds that stem from it, as Kant notes in the following passage: “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). Kant addresses this issue by introducing a distinction between two senses in which

---

44 Kant makes note of a possible contradiction in the concept of a propensity to evil when he writes: “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). Now an act of the human being is called a deed when it is caused by the free exercise of his will. If an act is not a deed, then it may not be imputed to the human being, and any evil that results from it would be a physical evil, but not a moral evil. If the propensity to evil is to be a moral evil, and not a physical evil, it would therefore have to be a deed. But a propensity is also a subjective determining ground of the power of choice that precedes every deed, and hence is itself not yet a deed, and so there can only be a propensity to physical evil, and not a moral one. Kant’s way out from this is to distinguish between two senses of the word “deed” in the following passage: “Now the term “deed” can in general apply just as well to the use of freedom through which the supreme maxim (either in favor of, or against, the law) is adopted in the power of choice, as the use by which the actions themselves (materially considered, i.e. as regards the objects of the power of choice) are performed in accordance with that maxim” (6:31). Both the intelligible deed through which the fundamental maxim is chosen and the empirical deeds that either obey or resist the moral law materially are deeds, and can therefore be imputed to the human being.
deed can be used. For a deed can either be *empirical*, or it can be *intelligible*. Empirical deeds are deeds that take place in time, and these are deeds which can resist the moral law materially as *peccatum derivatum*, that is, derivative sin (6:31). Radical evil, as the subjective determining ground of the power of choice, precedes every *empirical deed*, and hence cannot itself be an empirical deed, on pain of circularity. But empirical deeds are not the same as intelligible deeds. Empirical deeds occur in the empirical realm and intelligible deeds occur in the intelligible realm.

Insofar as radical evil involves the adoption of a fundamental maxim that is in opposition to the moral law by the free power of choice, it falls under the definition of a deed: “An action is called a *deed* insofar as it comes under obligatory laws and hence insofar as the subject, in doing it, is considered in terms of the freedom of his choice. By such an action the agent is regarded as the author of its effect, and this, together with the action itself, can be imputed to him…” (6:223). This means that if the adoption of radical evil is not a deed, then the human being cannot be regarded as the author of this effect, and the adoption of radical evil cannot be imputed to him. Therefore, insofar as the adoption of radical evil cannot be an empirical deed because it is supposed to precede every empirical deed, the adoption of radical evil has to be an intelligible deed because it can be imputed to us. And as an intelligible deed, it is the formal ground of every empirical deed contrary to law, leading Kant to call it the *peccatum originarium*, that is, original sin (6:31).

But the adoption of radical evil is more than just a deed that takes place in the intelligible realm. It is also a *moral evil*. Remember that for something to be a moral evil, it must involve a choice freely made by the *Willkür*. Now the Incorporation Thesis is defined by Kant in the

---

But when the concept of a propensity is understood as a subjective determining ground of the power of choice that precedes every deed, and hence is itself not yet a deed, it is understood as a subjective determining ground of the power of choice that precedes every *empirical* deed, and hence is itself not yet an *empirical* deed. This does not result in a contradiction because the propensity to evil is adopted in the power of choice through an *intelligible* deed, in which the human being’s fundamental maxim is determined to be against the law.
following manner: “... Freedom of the power of choice has the characteristic, entirely peculiar to it, that it cannot be determined to action through any incentive except so far as the human being has incorporated it into his maxim (has made it into a universal rule for himself, according to which he wills to conduct himself)” (6:24). This means that the presence of an incentive in a maxim is never accidental. The human being must incorporate the incentives into his maxim in order to act on them. Both the moral and nonmoral incentives end up being incorporated into the human being’s maxim because he wills it, insofar as the human being is so constituted to be receptive to both the moral law and to the impulses of his sensuous nature. As such, what makes a maxim good or evil depends on the Willkür’s choice to subordinate the moral incentive to the nonmoral incentive in its construction of the maxim, or vice versa. Kant writes: “… The difference, whether the human being is good or evil, must not lie in the difference between the incentives that he incorporates into his maxim (not the material of the maxim) but in their subordination (in the form of the maxim): which of the two he makes the condition of the other” (6:36). The maxim in which the nonmoral incentive is subordinated to the moral incentive is considered to be good. And the maxim in which the moral incentive is subordinated to the nonmoral incentive is considered to be evil. In both these cases, Kant is insistent that the morality of the deed involves an active choice on the part of the Willkür. For Kant, moral evil is not a simple failure to choose the moral incentive: the Willkür chooses “a real and opposite determination” to the moral incentive, the nonmoral incentive, and subordinates the moral incentive to it (6:22n).

This account becomes problematic when we try to apply it to the adoption of radical evil. The adoption of radical evil, remember, is a deed that takes place in the intelligible realm. But by placing this deed on the noumenal side of the noumenal-phenomena divide, Kant has put it beyond the reach of the appearances. This means that nothing from the phenomenal side can have any
influence on the will as it makes the choice to adopt radical evil. The way Morgan puts it, “The problem we face is that in imagining the will between morality and self-love, prior to its immersion in the world, we have pared away virtually every element of the self that might provide it with reasons – its desires, its emotions, its values, its membership in a society, its history, its individual identity” (Morgan 2005, p. 77). In other words, the nonmoral (sensuous) incentives that have their origins in the body cannot have any influence whatsoever on the choice being made in this intelligible deed, because they belong on the wrong side of the divide. But a moral evil is only produced when the Willkür chooses to prioritize the nonmoral incentive over the moral incentive in his maxim. This becomes impossible if there is no nonmoral incentive that the Willkür can subordinate the moral incentive to, because the only incentive available to the Willkür in the intelligible realm is the incentive of morality. By placing the adoption of the fundamental maxim on the noumenal side of the noumenal-phenomena divide, Kant has deprived this deed of the material conditions necessary for its being an evil deed. For nothing is available to the Willkür in the intelligible realm besides the moral incentive, and this means that the Willkür only has the moral incentive to choose from when it forms its fundamental maxim. Therefore, the Willkür can only choose to adopt a good fundamental maxim, not an evil fundamental maxim.

But even if, per impossibile, a nonmoral incentive could be made available to the Willkür for this intelligible deed, the adoption of radical evil remains problematic. By placing this deed on the noumenal side of the noumenal-phenomena divide, Kant has also put it beyond the laws of sensibility. As such, the only law that can determine this intelligible deed is the law of reason, as Kant notes in the Groundwork: “Since the concept of causality brings with it that of laws in accordance with which, by something that we call a cause, something else, namely an effect, must be posited, so freedom, although it is not a property of the will in accordance with natural laws, is
not for this reason lawless but must instead be a causality in accordance with immutable laws but of a special kind; for otherwise a free will would be an absurdity” (4:446). And a will that obeys the law of reason would decide to adopt a fundamental maxim that is morally good. For the law of reason is none other than the moral law. Conversely, a will that decides to adopt a fundamental maxim that is morally evil cannot be a will that is solely determined by the laws of reason. Insofar as a free will is not lawless, it follows that a will that is not determined by a law of reason must be a will that is determined by natural laws. Therefore, even if a nonmoral incentive was made available to the Willkür for this intelligible deed, there is no law of causality available to the Willkür that could make for the adoption of radical evil. Insole describes the problem in the following terms: “If the noumenal self is perfectly rational, it cannot do otherwise than the good, as ‘the determination according to natural law is abolished on account of freedom’; but, if the noumenal self can do otherwise than the good, then it would seem that it is not perfectly rational, in which case it must be subjected in some way to sensuous and irrational impulses, in which case it is not an atemporal first cause” (Insole 2013, p. 127).

From this, we can describe the problem that Kant created when he made the adoption of radical evil an intelligible deed thusly: given that a rational being would choose the good, and given that the noumenal free self is not subjected to any interfering sensuous impulses, it seems that a rational noumenal being would not choose to do otherwise than the good (Insole 2013, p. 127). Indeed, this can be put in even stronger terms: given that a rational being is determined by the laws of reason to choose the good, and given that the noumenal free self is given no alternative to the good in the form of interfering sensuous impulses, it seems that a rational noumenal being has no choice but to choose the good. But choosing the good in the context of this intelligible deed involves adopting a fundamental maxim that is morally good, and this is not what Kant thinks
actually happens – the doctrine of radical evil maintains that, in the context of this intelligible deed, every human being chooses to adopt a fundamental maxim that is morally evil. This makes no sense, given our considerations to this point. There aren’t any nonmoral incentives available to the Willkür that has to decide the human being’s fundamental maxim, and there is no law through which the Willkür, as a causality, is able to select anything other than a fundamental maxim that is good. Despite this, Kant expects us to believe that the human being exercises his will (as a causality) to select a fundamental maxim that is evil when the (intelligible) deed is done. This drives us to ask the following question: How is it possible to choose a fundamental maxim that is evil? And yet it must be possible to choose a fundamental maxim that is evil, because every human being actually does choose a fundamental maxim that is evil.

**Fallibility and Infallibility**

The problem we are faced with here is this: As a noumenal first cause, we have neither the material conditions (in the form of nonmoral incentives) nor the formal conditions (in the form of a law of causality) to choose a fundamental maxim that is evil. The will is always determined by the law of reason, to some degree. This is evident in Kant’s rejection of the diabolical will, which chooses to resist the moral incentive: “… An evil reason as it were (an absolutely evil will), would on the contrary contain too much [to provide a ground of evil in the human being], because resistance to the law would itself be thereby elevated to incentive (for without any incentive the power of choice cannot be determined), and so the subject would be made a diabolical being” (6:35). The diabolical will is evil not because it chooses to incorporate the nonmoral incentive

---

45 Kant’s denial of the diabolical will is controversial. It goes against the intuitions that we have that human beings can be motivated to action by the mere fact that the act is contrary to the moral law. As Augustine records in his Confessions: “A pear tree there was near our vineyard, laden with fruit, tempting neither for color nor taste. To shake and rob this, some lewd young fellows of us went, late one night (having according to our pestilent custom prolonged our sports in the streets till then), and took huge loads, not for our eating, but to fling to the very hogs,
into its maxims, nor because it chooses to subordinate the moral incentive to the nonmoral incentive in its maxims, but because it elevates resistance to the moral law into an incentive: “To think of oneself as a freely acting being, yet as exempted from the one law commensurate to such a being (the moral law), would amount to the thought of a cause operating without any law at all (for the determination according to natural law is abolished on account of freedom): and this is a contradiction” (6:35). Kant’s point here is the same as it was in the *Groundwork*: the will is a kind of causality, and “the concept of causality brings with it that of laws in accordance with which, by something that we call a cause, something else, namely an effect, must be posited” (4:446). As *free*, the causality of a diabolical will is considered to be independent from the determination of the laws of nature. And as *diabolical*, the diabolical will is considered independent from the only other law that could possibly determine it, the law of reason. The diabolical will is therefore a causality that is *lawless*, and this is absurd insofar as the concept of causality brings with it the concept of *laws* (4:446), which is why Kant describes the diabolical will as a contradiction.  

46 Caswell gives an alternative account of the contradiction of the diabolical will. He objects to the idea of a will that violates the moral law out of a love for evil as such, which would subordinate the moral incentive to the incentive to defy the moral law. This would result in a *Gesinnung* where “love of evil is the highest priority, and the condition under which all other pursuits are promoted: either self-love or else morality assume second place, leaving the other to occupy third place” (Caswell 2007, p. 151). But this is logically impossible. Caswell explains why in the following passage: “The problem is that it is impossible to subordinate any interest to an interest in its opposite. Consider a non-moral example. I may have competing interests in my career and in gold. Now it is possible for me to subordinate my career interest to my gold interest, neglecting professional duties in order to spend more time golfing; or, alternatively, I could make my job my higher priority, and play gold only when my work schedule allowed. But it is not possible to subordinate my love of golf to a hatred of golf. If I did hate golf, and thus had an interest in avoiding golf whenever possible, I could not pursue this interest in a way that accommodated my love of golf, or my interest in...
order to determine the difference between a will that has a good fundamental maxim in its heart and a will that has an evil fundamental maxim in its heart, we must compare them both in order to ascertain how the law of reason determines them (insofar as a will that is not determined by reason would be a diabolical will).

In the *Groundwork*, Kant introduces the concept of an imperative by making a distinction between two types of wills. Wills of the first type are *fallibly determined* by reason: “… If reason solely by itself does not adequately determine the will; if the will is exposed to subjective conditions (certain incentives) that are not always in accord with objective ones; in a word, if the will is not in itself completely in conformity with reason (as is actually the case with human beings), then actions that are cognized as objectively necessary are subjectively contingent, and the determination of such a will in conformity with objective laws is *necessitation*…” (4:412-413).

For Kant, to say that to do or to omit something would be good is to provide a *reason* or *incentive* to do or to omit something. And to say that to do or to omit something would be *subjectively* good is to say that the reason or incentive to do or to omit something is regarded by the subject as holding for his own will. By contrast, to say that to do or to omit something would be *objectively* good is to say that the reason or incentive to do or to omit something holds for the will of every rational being. In a will that is fallibly determined by reason, it is possible for these two notions of good to disagree. Although all rational wills may agree that to do or to omit something would be good, the human being may decide that it is against his interests to do or omit something, and choose not to act in accordance with the law of reason for prudential reasons. Such an action would be cognized playing gold whenever possible. These two interests (golf-love and golf-hatred) cannot share a home in the same will” (Caswell 2007, p. 152). In the same way, the moral law cannot be both an *incentive* and a *disincentive* for one and the same will. If the moral law in us were not an incentive of the power of choice, then there would be logical space for a diabolical will that treats the moral law as a disincentive. But the moral law is an incentive of the power of choice (6:22n). Therefore, the diabolical will that treats the moral law as a disincentive must at the same time treat the moral law as an incentive, and this is a contradiction.
as one that is objectively necessary but subjectively contingent (4:413). In such case, the moral law has to necessitate the fallible will to comply with the dictates of reason, and this necessitation takes the form of an imperative: “All imperatives are expressed by an ought and indicate by this the relation of an objective law of reason to a will that by its subjective condition is not necessarily determined by it (a necessitation). They say that to do or to omit something would be good, but they say it to a will that does not always do something just because it is represented to it that it would be good to do that thing” (4:413).

This may be compared to the wills of the second type, which are infallibly determined by reason: “If reason infallibly determines the will, the actions of such a being that are cognized as objectively necessary are also subjectively necessary, that is, the will is a capacity to choose only that which reason independently of inclination cognizes as practically necessary, that is, as good” (4:412). For a will that is infallibly determined by reason, there can be no disagreement between the two notions of good. If it is said that to do or to omit something would be objectively good, then this would be regarded as a reason or incentive for the subject to act, that is, as a reason or incentive holding for its own will as well. This means that the moral law does not need to compel the will that is infallibly determined to act according to its strictures, because such a will is by its nature necessarily obedient to it (4:413). Therefore, the moral law does not take the form of an imperative for the will that is infallibly determined: “A perfectly good will would, therefore, equally stand under objective laws (of the good), but it could not on this account be represented as necessitated to actions in conformity with law since of itself, by its subjective constitution, it can be determined only through the representation of the good. Hence no imperatives hold for the divine will and in general for a holy will: the “ought” is out of place here, because volition is of itself necessarily in accord with the law” (4:414). As the passage indicates, there are two types of
wills that are infallibly determined by reason. The divine will, which is the will belonging to an infinite being, is one example of a will that is infallibly determined by reason. The holy will is the other example of a will that is infallibly determined by reason, and Kant is almost certainly referring to the holiness of the finite will in this context, in contradistinction to the holiness of the infinite will (insofar as the divine will is trivially also a holy will). Thus, finitude is not the determinant as to whether reason fallibly determines the will, insofar as the will of a finite being can be fallibly determined by reason, and it can also be infallibly determined by reason.

Kant’s treatment of the divine will is found in his Lectures on the philosophical doctrine of religion. There he argues that God is free, despite not having the power to do otherwise. Hence:

The human being, however, can always decide something else, e.g. a human being, instead of being benevolent in this case, could also not be that. But it is precisely this which is a lack of freedom in the human being, since he does not always act according to his reason; but in God it is not due to the necessity of his nature that he can decide only as he does, but rather it is true freedom in God that he decides only what is in conformity with his highest understanding (28:1068).

What is the distinction between an infinite being and a finite being? We know what an infinite being is for Kant: It is an ens realissimum. Finite beings, then, are any beings that are not ens realissimum. Now an ens realissimum is the most real being, and it is distinguished from finite beings that have only some reality, which is what Kant does in the following passage: “Every thing must have something positive which expresses some being in it. A mere not-being cannot constitute any thing. The concept de ente modo negativo is a concept of a non entis. Consequently, since each thing must have reality, we can represent every possible thing either as an ens realissimum or as an ens partim reale, partim negativum. But in the same case of any thing which has only some reality, something is still always lacking, and hence it is not a complete thing. A highest thing, therefore, would have to be one which has all reality. For in this one case I do have a thing whose thoroughgoing determination is bound up with its concept, because it is thoroughly and completely determined with respect to all possible praedicates opposites. Consequently, the concept of an ens realissimum is the very concept of an ens summum; for all things except this being are partim reale, partim negativum and just because of this their concepts are not thoroughly determined” (28:1013-1014). Therefore, a finite being can be distinguished from an infinite being in two ways with respect to its reality: either it lacks some reality that the ens realissimum has, or the reality that it shares with the ens realissimum is limited in some way. This is what it means to be partly real and partly negative (ens partim reale, partim negativum): “… Thus if every negative concept is derivative in that it always presupposes a reality, then every thing in its thoroughgoing determination as an ens partim reale, partim negativum also presupposes an ens realissimum with respect to its realities and negations, because they are nothing but limitations of the highest reality. For when I entirely remove some realities from the concept of an ens realissimum, there arise negations which give me the concept of an ens partim reale, partim negativum when I combine them with the remaining realities: hence the concept of an ens realissimum contains simultaneously the ground for every other concept” (28:1014). It is this insight that forms the basis of Kant’s proof in The only possible argument in support of a demonstration of the existence of God (28:1015).
There are, therefore, no subjective conditions (nonmoral incentives) in God that are opposed to the objective conditions (moral incentives) in him. And evil is therefore a logical impossibility for God. For moral evil is only possible as “a real and opposite determination of the power of choice”, that is, as a real opposition, which is logically inconceivable in a perfect being, as Kant notes: “… In the most real being of all there cannot be any real opposition or positive conflict among its own determinations, for the consequence would be a deprivation or a lack, and that would contradict its supreme reality” (2:86). In contrast to the divine will, the finite holy will is incapable of evil due to its being a real impossibility for the finite holy will. The discussion of the finite holy will takes place in the Metaphysics of Morals, where Kant writes: “For finite holy beings (who could never be tempted to violate duty) there would be no doctrine of virtue but only a doctrine of morals, since the latter is autonomy of practical reason whereas the former is also autocracy of practical reason, that is, it involves consciousness of the capacity to master one’s inclinations when they rebel against the law…” (6:383). This is not a case where the subjective conditions (nonmoral incentives) are always in accord with the objective ones (moral incentives), like in the divine will. The finite holy will can experience temptation. But there is no level of temptation that is able to overcome the finite holy will’s commitment to the moral law. Thus, evil is logically possible for the finite holy will, insofar as nothing is being simultaneously affirmed and denied of the very

---

48 As Chignell notes, Kant never abandoned his commitment to the deductive validity of the Pre-Critical argument given in The only possible argument in support of a demonstration of the existence of God (1763). This is in spite of the criticisms he made to the traditional arguments for the existence of God in the Transcendental Dialectic. In the Lectures on the philosophical doctrine of religion, Kant refers to the Pre-Critical argument when he writes: “… This proof is not apodictically certain; for it cannot establish the objective necessity of an original being, but establishes only the subjective necessity of assuming such a being. But this proof can in no way be refuted, because it has its ground in the nature of human reason. For my reason makes it absolutely necessary for me to assume a being which is the ground of everything possible, because otherwise I would be unable to know what in general the possibility of something consists in” (28:1034). And in a Reflexion dated circa 1785, Kant again refers to the Pre-Critical argument when he writes: “The possibilities of things, which can only be regarded as determinations of a single universal possibility, namely of the highest being, prove the existence of the realissimi as a sum-total, consequently, if understanding is a reality, they prove that it is intelligent” (R6278). These passages are evidence that Kant continued to conceive God as the ground of all realities (the ens realissimum) well into the Critical period of his thought.
same thing here. But it is really impossible for the finite holy will, insofar as the good tendency in him is always sufficient to cancel out the evil tendency in him.

It is the subjective constitution of the will that determines whether the will is fallibly or infallibly related to reason. If reason does not adequately determine the will, then when the fallible will is exposed to nonmoral incentives, it might not do something just because it has been represented to it that it would be good to do that thing (4:413). But if reason adequately determines the will, then even though the infallible will is exposed to nonmoral incentives, they have no influence on the will whatsoever, because the will chooses only that which reason independently of inclination cognizes as practically necessary (4:412). But is the subjective constitution of the will an intrinsic property or an extrinsic one? Intrinsic properties are properties that a subject can have even if it is lonely, that is, if it is the only thing in existence.49 Here we find Langton discussing intrinsic properties with respect to substances: “A substance is a thing which can exist absolutely, ...

---

49 The concept of loneliness is introduced alongside that of accompaniment in the article “Psychophysical Supervenience”, where Kim provides the following definition: $G$ is rooted outside the objects that have it $= \text{def.}$ Necessarily any object $x$ has $G$ only if some contingent object wholly distinct from $x$ exists (Kim 1981, p. 60). If $G$ is rooted outside the objects that have it, then some contingent object wholly different $x$ has to exist in order for $x$ to have $G$. $G$ would therefore entail that $x$ must be accompanied. If $G$ is not rooted outside the objects that have it, then no contingent object wholly distinct from $x$ has to exist in order for $x$ to have $G$. $G$ would therefore be compatible with $x$ being lonely. Lewis has objected to this identification of intrinsicsness with loneliness in the following way: “Loneliness is just as extrinsic as accompaniment, yet certainly it does not imply accompaniment and certainly it is compatible with itself. If something is lonely – the cosmos, or some lesser otherworldly thing – its loneliness remains unrooted” (Lewis 1983, p. 199). This is because an object’s being accompanied or lonely both depend on whether something else exists, and whether a contingent object wholly distinct from $x$ exists or not is, by definition, not up to $x$. Therefore, accompaniment and loneliness are both extrinsic properties. This was followed by a later proposal in which Langton and Lewis defined a property as intrinsic iff the following four conditions were met:

1. A lonely thing can have the property $G$
2. A lonely thing can lack the property $G$
3. An accompanied thing can have the property $G$
4. An accompanied thing can have the property $G$

If these four conditions are met, then having or lacking $G$ is independent of loneliness or accompaniment, and Langton and Lewis define intrinsicsness as a property that is independent in this way (Langton-Lewis 1998, p. 334).
independently of all its relations to other things. A substance is the kind of thing that can exist on its own it can exist and be lonely. But nothing can exist without having properties. If a substance can exist on its own, it must have properties that are compatible with existing on its own. If a substance can be lonely, it must have properties compatible with loneliness. So a substance must have intrinsic properties” (Langton 1998, p. 19). This discussion is applicable to our consideration of the subjective constitution of the will. For the subjective constitution of the will consists of the properties that make up the will of the subject. If these properties were absent, there would be no will, no reason, and no (mental) subject. The subjective constitution of the will must therefore be an intrinsic property (or a collection of intrinsic properties). The nonmoral incentives, by contrast, have their origins in nature. If the phenomenal world were to suddenly cease to exist, these nonmoral incentives would disappear along with their causes. All that would remain are the intrinsic properties that make up the subject, the subjective constitution of his will, in its relation to reason, be it fallible or infallible.

**Vice and Virtue**

In the *Metaphysics of Morals*, Kant defines virtue in the following manner: “Now the capacity and considered resolve to withstand a strong but unjust opponent is fortitude (*fortitudo*) and, with respect to what opposes the moral disposition within us, virtue (*virtus*, *fortitudo moralis*)” (6:380). This definition is given as part of his discussion of duty, which is the

---

50 It should be observed that Kant’s concept of virtue closely corresponds to the Scholastic concepts of continence and fortitude. Aquinas describes continence in the following way: “Continence denotes, by its very name, a certain curbing, in so far as a man contains himself from following his passions. Hence continence is properly said in reference to those passions which urge a man towards the pursuit of something, wherein it is praiseworthy that reason should withhold man from pursuing; whereas it is not properly about those passions, such as fear and the like, which denote some kind of withdrawal: since in these it is praiseworthy to remain firm in pursuing what reason dictates, as stated above. Now it is to be observed that natural inclinations are the principles of all supervening inclinations, as stated above. Wherefore the more they follow the inclination of nature, the more strongly do the passions urge the pursuance of an object” (ST II-II, Q 155, Art. 2). For Aquinas, there is a distinction between the
necessitation of a will by the objective law of reason. According to Kant, necessitation here is a kind of constraint, and he divides these constraints into external constraints and internal constraints. External constraints may be imposed on us by others, and the duties that can be coerced in this way form the basis of Kant’s doctrine of rights (6:383). Internal constraints, by contrast, cannot involve coercion by others. The duties that form the basis of Kant’s doctrine of virtue are based on free self-constraint (6:383). Insofar as the impulses of nature that create obstacles to the moral disposition within us can only be resisted through self-constraint, it follows that “virtue” and the capacity for “self-constraint” are one and the same thing. Insofar as virtue is a capacity, however, we may understand it in terms of a counterfactual conditional: If the impulses of nature create obstacles to the moral disposition, the virtuous person would withstand it. This does not passions which urge a man towards the pursuit of something and the passions which urge a man to withdraw from something. Continence curbs the desires that seek to tempt the human being away from reason, and therefore pertains only to the former. Whereas it takes fortitude to endure and withstand those things to which we have an aversion, under which the latter falls: “... It belongs to fortitude to remove any obstacle that withdraws the will from following reason. Now to be withdrawn from something difficult belongs to the notion of fear, which denotes withdrawal from an evil that entails difficulty, as stated above in the treatise on passions. Hence fortitude is chiefly about fear of difficult things, which can withdraw the will from following reason” (ST II-II, Q 123, Art. 3). By contrast, Kant considers the faculty of desire to be responsible for the formation of both desire, which is connected to the feeling of pleasure, and aversion, which is connected to the feeling of displeasure (6:211). Insofar as inclination is nothing more than a habitual desire, we might call a disinclination a habitual aversion (6:212). Both of these can be obstacles to the fulfillment of one’s duty, as Kant describes in the following passage: “Impulses of nature, accordingly, involve obstacles within the human being’s mind to his fulfillment of duty and (sometimes powerful) forces opposing it, which he must judge that he is capable of resisting and conquering by reason not at some time in the future but at once (the moment he thinks of duty)” (6:380). Insofar as virtue is the capacity for resisting and conquering by reason these impulses of nature, it represents the moral strength to dutifully refrain from what we are naturally inclined to do, and dutifully act what we are naturally inclined to avoid.

The duties that fall under the doctrine of right pertain to our actions. These duties tell us what actions are obligated or prohibited by the moral law. It is therefore possible for an external authority to compel us to act in accordance with these duties. The duties that fall under the doctrine of virtue, by contrast, pertain to the incentives from which we act and the end for which we act. This setting of ends is internal and performs according to our sole discretion. Hence Kant writes: “... Determination to an end is the only determination of choice the very concept of which excludes the possibility of constraint through natural means by the choice of another. Another can indeed coerce me to do something that is not my end (but only a means to another’s end), but not to make this my end; and yet I can have no end without making it an end for myself. To have an end that I have not myself made an end is self-contradictory, an act of freedom that is not yet free” (6:381). That there is an end that is also a duty follows from the concept of the categorical imperative (6:385). The way Louden explains it is so: “We must assume that there are morally necessary ends, for if we don’t, “this will do away with all moral philosophy.” [Kant’s] reasoning is that if all ends are contingent, then all imperatives become hypothetical. If we are free to accept or reject any goal put before us whenever we are so inclined, then all commands prescribing maxims for actions are likewise open to rejection once the goal is dismissed. In other words (by contraposition), if there is a categorical imperative, there must be at least one morally necessary end” (Louden 1986, p. 482).
imply that there actually are impulses of nature in the human being that create obstacles to the moral disposition in him. That is, it is possible for a person to have the capacity to withstand the impulses of nature, and thereby be virtuous, even if he has no impulses of this kind, or even if the impulses in him pose no obstacle to his moral disposition. It is a matter of fact, not a matter of necessity, that every virtuous person has impulses of nature that create obstacles to the moral disposition.52

The concept of virtue as moral fortitude is opposed by two distinct concepts, as evidenced by what Kant writes in the following passage: “Virtue = +a is opposed to negative lack of virtue (moral weakness = 0) as its logical opposite (contradictione oppositum); but it is opposed to vice = +a as its real opposite (contrarie s. realiter oppositum)” (6:384). This distinction between logical opposition and real opposition can be traced back to Kant’s Pre-Critical writings.53 In Negative

---

52 This fact about virtue needs to be remembered, given Kant’s definition of virtue in a later passage: “Virtue is the strength of a human being’s maxims in fulfilling his duty. – Strength of any kind can be recognized only by the obstacles it can overcome, and in the case of virtue these obstacles are the natural inclinations, which can come into conflict with the human being’s moral resolution...” (6:394). Now the strength of a human being’s maxims in fulfilling his duty is a capacity. It is a property that a human being has (or can have). This is distinct from the recognition of strength, or the determination of whether a human being actually has this property. Therefore, in saying that strength of any kind can be recognized only by the obstacles it can overcome, Kant is indicating the epistemological role of natural inclinations in revealing an underlying metaphysical fact, the moral fortitude of a human being. This is precisely the point Kant makes in a later passage in the Metaphysics of Morals: “Virtue so shines as an ideal that it seems, by human standards, to eclipse holiness itself, which is never tempted to break the law. Nevertheless, this is an illusion arising from the fact that, having no way to measure the degree of a strength except by the magnitude of the obstacles it could overcome (in us, these are inclinations), we are led to mistake the subjective conditions by which we assess the magnitude for the objective conditions of the magnitude itself” (6:396-397).

53 Kant discusses this in the Religion in the context of noumenal and phenomenal virtue. He describes phenomenal virtue in the following terms: “When the firm resolve to comply with one’s duty has become a habit, it is called virtue also in a legal sense, in its empirical character (virtues phaenomenon). Virtue here has the abiding maxim of lawful actions, no matter whence one draws the incentives that the power of choice needs for such actions. Virtue, in this sense, is accordingly acquired little by little, and to some it means a long habituation (in the observance of the law), in virtue of which a human being, through gradual reformation of conduct and consolidation of his maxims, passes from a propensity to Vice to its opposite” (6:47). In its empirical character, virtue is opposed by vice, which we know from the Metaphysics of Morals consists in a real opposition to virtue (6:384). Insofar as vice is a material resistance to the moral law, it represents the desires and incentives that “hinder the human heart in complying with the adopted maxims” (6:29). If it has the moral strength to withstand these temptations, and comply with the maxims that it adopts, these maxims become abiding. But Kant also describes noumenal virtue, and he does so in the following terms: “However, that a human being should become not merely legally good, but morally good (pleasing to God) i.e. virtuous according to the intelligible character (virtus noumenon) and thus in need of no other incentive to recognize a duty except the representation of duty itself – that, so long as the foundation of the maxims of the human being remains impure, cannot be effected through gradual reform but must rather be effected through a revolution in the
Magnitudes, Kant explains an opposition between two things as one thing canceling out what is posited by the other (2:171). He then proceeds to explain logical opposition in the following passage: “... The opposition consists in the fact that something is simultaneously affirmed and denied of the very same thing. The consequence of the logical conjunction is nothing at all (nihil negativum irrepraesentabile), as the law of contradiction asserts. A body which is in motion is something; a body which is not in motion is also something (cogitabile); but a body which is both in motion and also, in the very same sense, not in motion, is nothing at all” (2:171). In the case of virtue, the logical opposition involves a positive predicate (moral fortitude) and a negative predicate (the absence of moral fortitude – moral weakness). But moral fortitude admits to degrees of perfection. For a person in whom moral fortitude is imperfectly present, the degree by which he falls short of perfect moral fortitude is the degree to which he is morally weak. It would therefore

Note that Chignell believes that real opposition can be divided into real oppositions that are subject-cancelling and real oppositions that are predicate-cancelling. A real opposition that is predicate-cancelling occurs when the opposition between two predicates cancels out each of their respective effects. The standard example of this type of real opposition is a ship that is subjected to a wind blowing east and a current flowing west. The effect of these two forces on the ship cancel each other out, and the ship is moved neither to the east nor to the west. There is, however, another type of real opposition, which Chignell describes using the following example: “Suppose A is the concept of a, and that A contains the predicates being water, and being XYZ (where “XYZ” refers, as usual, to some chemical compound other than H2O). Most philosophers would agree that these predicates are not logically opposed to one another. But they are really opposed when co-instantiated at a time; thus, in A they will “cancel out” not just one another but also the real possibility of a as a whole. In other words, any joint and simultaneous instantiation of these two predicates makes their bearer a really impossible being” (Chignell 2009, p. 173). The real opposition described here is subject-cancelling. It consists in a conflict between the nature of a thing and “a predicate which is not metaphysically compatible with that nature” (Chignell 2009, p. 173). And despite consisting in two positive predicates, the result of this real opposition is not the cancellation of the predicates (and their effects), but that of the subject, like in a logical opposition.

To consider another example, consider the predicates “light” and “dark”. Now it is evident that darkness is the absence of light. The opposition between light and dark is therefore an opposition between a positive predicate and its negation (a negative predicate), and not an opposition between two different positive predicates. Therefore, the opposition between light and dark is a logical opposition and not a real opposition. But it does not follow from this that the presence of these two predicates in the same subject is a contradiction (subject-cancelling). The presence of
be right to say that such a person is morally weak, even though this does not mean he is completely lacking in moral fortitude. Thus, what makes this a logical opposition is that a positive predicate (virtue) is opposed to a negative predicate (the absence of virtue), and not that the opposition is subject-cancelling (2:77).

This may be contrasted with real opposition, which Kant explains in the following passage: “The second opposition, namely real opposition, is that where two predicates of a thing are opposed to each other, but not through the law of contradiction. Here, too, one thing cancels out that which is posited by the other; but the consequence is something (cogitabile)” (2:171). In a real opposition, the two predicates are opposed to each other, and both these predicates are simultaneously possible in one body. Kant oftentimes uses opposing motive forces as his paradigmatic example of a real opposition: “The motive force of a body in one direction and an equal tendency of the same body in the opposite direction do not contradict each other; as predicates, they are simultaneously possible in one body. The consequence of such an opposition is rest, which is something (repraesentabile)” (2:171). But if one of these motive forces were to suddenly be removed, the body would no longer be at rest. The resting state of the body is

---

these two predicates in the same subject is a contradiction when their magnitudes are inconsistent. If we say of a room that it is both completely illuminated (light = 1) and completely dark (light = 0), then the result would be a contradiction. But if we say of a room that it is dimly-lit (light = 0.5), then there would be no contradiction, because even though there is light in the room, there is also darkness to the extent that the room is not completely illuminated, and the same can be said of the darkness in the room; mutatis mutandis. Hence, Kant writes in his Lectures on the philosophical doctrine of religion: “If one speaks of “mixed realities”, one is using an improper expression. For a mixture of a reality and a negation, of something and nothing, cannot be thought. If I am to mix something with something else, then I must have something actual; but negations are mere deficiencies. Hence if a thing has something negative along with what is real (for example, a darkened room, etc.) then in this case there is no mixing in of the negation, but rather a limitation of reality. Thus in the case cited I could not mix the negation darkness in with the light as something real, but rather the negative darkness arose when I reduced and limited the reality light. But the logical mixture of concepts is something wholly different. Here I can certainly say that the concept of a negation is mixed in with my concept of reality, for my concept of something negative is a concept every bit as much as my concept of something real is a concept” (28:1015). Therefore, the distinction between a logical opposition and a real opposition does not consist in the fact that the opposition is subject-cancelling or predicate-cancelling, because both logical opposition and real opposition can be either. It consists in the fact that a logical opposition is an opposition between a positive predicate and its negation (a negative predicate), and a real opposition is an opposition between two different positive predicates.
maintained only because the motive force of a body in one direction has its effect (movement) cancelled out by an equal tendency of the same body in the opposite direction. Analogously, the real opposition between virtue and vice may be thought of in terms of motive forces also. The moral law provides a person with an incentive to act in a certain way. The impulses of nature might provide a person with an incentive to not act in that way. As human beings, both these incentives are always simultaneously present in a person (6:36). If the moral incentive does not have the strength to overcome the nonmoral incentive, then the consequence of such an opposition is rest (inaction), even though the human being has already made it his maxim to act in a certain way, because “between the maxim and deed there still is a wide gap” (6:47). This is what happens in frailty, when the “good, which is an irresistible incentive objectively or ideally (in thesi), is subjectively (in hypothesi) the weaker (in comparison with inclination) whenever the maxim is to be followed” (6:29). But if one of these motive forces were to suddenly be removed, the person would no longer be inactive. The inaction of the person is maintained only because he does not have the fortitude to overcome the obstacles that prevent him from carrying out his maxim.56

56 Let us take a moment and consider how divine grace would look like under this account of radical evil. In this chapter, radical evil is portrayed as a choice that enfeebles the moral character of the human being. The resulting absence of moral strength leaves the human being impaired in the action of his moral duties. The role of divine grace is as assistance to the human being in the execution these moral duties. In particular, God provides the human being with the moral strength that is lacking: “… Granted that some supernatural cooperation is also needed to his becoming good or better, whether this cooperation only consist in the diminution of obstacles or be also a positive assurance, the human being must nonetheless make himself antecedently worthy of receiving it; and he must accept this help (which is no small matter), i.e. he must incorporate this positive increase of force into his maxim” (6:44). Per Mariña’s discussion of divine grace, Kant has two notions of grace besides his general understanding of grace as God’s unmerited favor (Mariña 1997, p. 385). She writes: “The first of these corresponds to a kind of grace that Kant believes cannot be brought into our practical maxim. It has to do with God’s supernatural cooperation with our becoming better persons, that is, with how God may affect the will itself such that its very desires and motives will become different” (Mariña 1997, p. 385). And then: “The second notion corresponds to the kind of divine aid which must be laid hold of by the person. It differs from the first in that such aid does not alter a person’s will at the outset, but is, rather, a some historical occurrence – a person or situation – to which the person must respond in some way. Only after the practical and existential import of the person or situation has been assessed and interiorized by the individual can it affect a person’s character” (Mariña 1997, p. 386). It is the second of these two conceptions of grace that is being described in the 6:44 passage. The human being must make himself worthy of this divine assistance, and only afterwards would this positive increase in moral strength be made available to him.
Now if we are going to be precise, it must be noted that the concept of virtue does not map perfectly onto the concept of moral fortitude (or moral strength). Kant makes this clear in the following passage: “Virtue signifies a moral strength of the will. But this does not exhaust the concept; for such strength can also belong to a holy (superhuman) being, in whom no hindering impulses would impede the law of its will and who would thus gladly do everything in conformity with the law” (6:405). This is because the concept of duty is inapplicable to a holy being.

Remember that Kant defines an imperative in the *Groundwork* as the relation of an objective law of reason to a will that by its subjective constitution is not necessarily determined by it (4:413). He then proceeds to make an exception for the divine will and the holy will, saying: “… The “ought” is out of place here, because volition is of itself necessarily in accord with the law. Therefore imperatives are only formulae expressing the relation of objective laws of volition in general to the subjective imperfection of the will of this or that rational being, for example, of the human will” (4:414). Imperatives hold for the human being because the subjective constitution of his will does not necessarily conform with the objective law of reason. Even though the imperative says that to do or to omit something would be good, it says it to a will that may not do something just because it is represented to it that it would be good to do that thing (4:413). But for a holy being, in whom no hindering impulses would impede the law of its will and who would thus gladly do everything in conformity with the law, there can be no imperatives. For this reason, the concept of duty, understood as an action that someone is obligated to perform by the categorical imperative (6:223), is inapplicable to holy beings. But the concept of virtue is inextricably linked to that of duty in the following passage: “Virtue is, therefore, the moral strength of a human being’s will in fulfilling his duty, a moral constraint through his own lawgiving reason, insofar as this constitutes itself an authority executing the law” (6:405). The holy being does not encounter hindering
impulses that would impede the law of its will like the human being does. Ipso facto, the holy being does not have a duty to fulfill, and virtue, as the moral strength needed to fulfill one’s duty, would be absent in a holy being.

Yet this does not mean that a holy being does not have moral fortitude (or moral strength). For the opposition to moral strength can be divided into its logical opposite and its real opposite. If we assume that a holy being does not have moral strength because there are no natural impulses in him that can prevent his will from acting in conformity with the moral law, then we have effectively ascribed to this holy being the logical opposite of moral strength, moral weakness. For a being is morally weak to the extent that his moral strength is lacking. But this is absurd, because a holy being in whom moral strength is absent would be neither holy nor superhuman. Insofar as a holy being does not have hindering impulses that can prevent his will from acting in conformity with the moral law, there is nothing that opposes the moral law in such a being. From this, what follows is that moral strength does not have a real opposite in a holy being. Therefore, the moral strength of a holy being has neither a logical opposite nor a real opposite. But this connection is merely contingent, because it is possible to imagine a being whose moral strength has a logical opposite, but not a real opposite. Imagine a being who is lucky—things have turned out for him such that there are no hindering impulses that are obstacles to his will acting in conformity with the moral law. The moral strength of a lucky being would therefore be without a real opposite as a matter of fact. From this, however, we are not licensed to infer that the moral strength of a lucky being has no logical opposite, because it is possible for such a being to be morally weak, but have the good fortune of living his entire life without ever being faced with hard moral decisions. The difference between the two cases lies in their attitude to the following counterfactual conditional: If he encounters a hindering impulse in him, he would still act in conformity with the law. It is
impossible for any hindering impulse to exist in a holy being, and this makes the counterfactual conditional true for a being that is holy. But it is possible for a hindering impulse to exist in a being who has had the good fortune to avoid them all his life, and he might not (or would not) act in conformity with the moral law in the face of these hindering impulses, and this makes the counterfactual conditional false for the being that was only lucky. His moral strength was lacking, but because he had the good fortune to have avoided all hindering impulses, this moral weakness never had the occasion to be revealed.

Given this analysis, the opposition between a fallible and infallible will should be understood in terms of moral strength and its privative opposite, moral weakness.\(^57\) For when a will is infallibly determined by reason, it is so determined by the moral strength of the individual. But a will that is fallibly determined by reason does not imply that the individual is entirely without moral strength. Insofar as his will is sometimes determined by reason, it is to that extent that he does have moral strength. And when reason fails to determine his will, and he does not act as he

\(^{57}\) In the traditional privation theory of evil, evil is understood as a privation of the good. It is an absence of a goodness a thing ought to have. Hence we find Aquinas providing the following explanation of evil: “Now, evil is in a substance because something which it was originally to have, and which it ought to have, is lacking in it. Thus, if a man had no wings, that is not an evil for him, because he was not born to have them; even if a man does not have blond hair, that is not an evil, for, though he may have such hair, it is not something that is necessarily due to him. But it is an evil if he has no hands, for these he is born to, and should, have – if he is to be perfect. Yet this defect is not an evil for a bird. Every privation, if taken properly and strictly, is that which one is born to have, and should have” (SCG 3.1.6.1). The traditional privation theory of evil is predicated on the Aristotelian metaphysics of natures and natural ends. It is by reference to the nature of a thing that we are able to say, concerning a good that is absent, that it ought to be there, and its absence is a privation. In the case of sin, the will that ought to be ordered towards a higher good, abandons this higher good for some inferior good instead. The absence of this right order is the privation that constitutes sin: “Now there is a certain order of these various things that are man’s goods, based on the fact that what is less primary is subordinated to what is more primary. Hence, a sin occurs in our will when, failing to observe this order, we desire what is only relatively good for us, in opposition to what is absolutely good” (SCG 3.2.108.6). The parallel to Kant is worth noting here. For Kant describes evil in much the same way: “Hence the difference, whether the human being is good or evil, must not lie in the difference between the incentives that he incorporates into his maxim (not in the material of the maxim) but in their subordination (in the form of the maxim): which of the two he makes the condition of the other. It follows that the human being (even the best) is evil only because he reverses the moral order of his incentives in incorporating them into his maxims” (6:36). The human being ought to order his incentives in the right way, by subordinating his nonmoral incentives to the moral incentive, and when he fails to do so, the order that ought to be there, is not, Kant’s absence of right order is therefore a privation, and because evil just is this reversal of the moral order of incentives, it follows that Kantian evil is a privation.
knows he should, to that extent is his moral strength lacking. It is not a power that grounds this deviation from the moral law, but rather the absence of power, a moral weakness. To borrow an analogous example from Wood: “An indifferent swimmer may have the power to save himself if he falls into deep water. But he has no power, only a possibility, of drowning in the same eventuality. Although he has the power to swim, he may drown if he does not exercise the power effectively, due (say) to confusion or panic. In the latter case, his possibility of drowning is due to a kind of weakness or lack of power, though not to a lack of power to swim” (Wood 1984, pp. 81-82). Likewise, a fallible will has the power to save itself in the face of temptation. All that it needs to do is maintain its commitment to the moral law, and stand strong against the temptation. But there is no opposing power that grounds the possibility where the will succumbs to temptation. This possibility comes to be due to a lack of power (moral weakness). Although the moral strength is there, and is sufficient to overcome the temptation, the fallible will may yet succumb to it if this power is not exercised effectively, due (for example) to a wavering commitment to the moral law. Put another way: “Not every possibility is a power. Some possibilities, in fact, are due to a lack of power” (Wood 1984, p. 81). So we may surmise that it is only fortitude (moral strength) that is something real and positive, whereas moral weakness is a negation of moral strength, a lack of fortitude, a privation of its power. And because moral weakness is derivative of moral strength in this way, it cannot exist without moral strength, even though moral strength can perfectly well exist without moral weakness, as in the case of a holy being.

Frailty and Embodiment

In the *Groundwork*, Kant introduces the idea that there is a natural dialectic between happiness and morality. The human being finds in himself a moral cognition, which he uses for the practical appraisal of all moral matters: “Inexperienced in the course of the world, incapable
of being prepared for whatever might come to pass in it, I ask myself only: can you also will that your maxim become a universal law? If not, then it is to be repudiated, and that not because of a disadvantage to you or even to others forthcoming from it but because it cannot fit as a principle into a possible giving of universal law, for which lawgiving reason, however, forces from me immediate respect” (4:403). This moral cognition is present in the common human understanding, and although, Kant writes, it is not represented in the familiar formulae of the categorical imperative, it is what the human being “actually has always before its eyes and uses as the norm for its appraisals” (4:403-404). But this moral cognition does not have full run over the human being unopposed. For Kant observes that, “The human being feels within himself a powerful counterweight to all the commands of duty, which reason represents to him as so deserving of the highest respect – the counterweight of his needs and inclinations, the entire satisfaction of which he sums up under the name happiness” (4:405). Morality and happiness are both necessary components of the natural dialectic. The conflict that the natural dialectic represents simply would not be able to exist without either. Insofar as the human being is a creature that belongs to both the sensible world and the intelligible world, he finds himself trapped in this conflict between his duties and his inclinations, pulled in one direction by the demands of the moral law in him and pulled in the other direction by the claims happiness makes on him.

Michalson traces these needs and inclinations to the fact of man’s embodiment. The reason why radical evil is universal is explained, according to him, by this: “… The connecting thread is provided by the body – every moral agent has one. What is more, the definitive feature of a moral agent is the struggle to subordinate the incentives that emerge because we are embodied to the incentive arising out of reason” (Michalson 1990, p. 69). It is because human beings are embodied that they have the needs and inclinations that they do. If not so embodied, the needs and
inclinations that form this powerful counterweight to the commands of duty would be gone as well, and there would be nothing that stands in the way of the human being’s performance of his duty. For this reason, Michalson makes the observation that “moral evil arises out of the competition for control of character between sensuous and rational incentives: rationality would have no competition if it were not embodied” (Michalson 1990, p. 69). Even so, Michalson is careful not to blame the body for the moral evil in human beings, because it has to be the Willkür that decides to give in to the temptations arising from the human being’s needs and inclinations, instead of obeying the commands of duty given to him by reason. For otherwise the moral evil in human beings could not be imputed to him, and this evil would thereby cease to be a moral evil, and become instead a physical evil. The point that emerges in Michalson’s discussion of embodiment is that, even though it is the free Willkür that goes wrong, it is the body that provides the free Willkür with the opportunity to go wrong (Michalson 1990, p. 69). This is his explanation for the

---

58 Kant is unambiguous in his insistence that the moral evil of the human being is not to be blamed on his natural inclinations (or the body from which they arise). This is evident in his criticism of the Stoics: “… However, those valiant men mistook their enemy, who is not to be sought in the natural inclinations, which merely lack discipline and openly display themselves unconcealed to everyone’s consciousness, but is rather as it were the invisible enemy, one who hides behind reason and hence all the more dangerous” (6:57). There are two factors behind this insistence. First, the incorporation thesis dictates that the power of choice cannot be determined to action through any incentive except so far as the human being has incorporated it into his maxim (6:24). This would not be the case if the natural inclinations were to blame for the human being’s moral evil. For it would mean that the nonmoral incentives supplied by these natural inclinations were by themselves sufficient to determine the human being’s power of choice to action, without any preceding act of incorporation. This would make the human being’s power of choice an arbitrium brutum, and in the absence of freedom, there can be no moral responsibility. Second, the natural inclinations have their origins in the human being’s predispositions to animality and humanity. The predispositions to animality is the source of the human being’s mechanical self-love, from which his instincts for self-preservation, the propagation of the species, and community with other human beings all originate (6:26). The predisposition to humanity is the source of the human being’s comparative self-love, without which he would never have been able to form the idea of happiness, as Wood explains: “Happiness is an idea we make for ourselves through imagination and reason. Put otherwise, the desire for it is a second-order desire for the satisfaction of a certain rationally selected set of first-order inclinations. It is far from being the case that every finite being of needs would be able to form such an idea; nor, even if it did form the idea, would every such being have a reason to desire its object” (Wood 1998, p. 145). Neither the predisposition to animality nor the predisposition to humanity are evil. In fact, Kant says the precise opposite they are both good. Hence: “All these predispositions in the human being 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). And neither of these would be true if the natural inclinations arising from our original predispositions to animality and humanity were to blame for the moral evil of human beings.
“latent resentment” of the body that we find throughout the Religion: “There is, after all, a profound connection between transcendental method and the “disembodied” standpoint – as though we “get it right”; philosophically, only to the extent that we bracket (if not escape altogether) the conditions of our embodiment, as well as the effects” (Michelson 1990, p. 69).

Michelson is exactly right on this point: rationality would have no competition if it were not embodied. The decision we make in the noumenal realm determines our commitment to the moral law. If we are committed to the moral law, then we will have the moral strength of will to be unwavering in our duty, even when we are faced with obstacles. But if we are not committed to the moral law, then our moral weakness of will becomes exposed when we are faced with obstacles, and we end up wavering in our duty. These obstacles are not present in the noumenal realm, and so the strength of our commitment to the moral law cannot be determined there, because we cannot measure the degree of a strength except by the magnitude of the obstacles it could overcome (6:396). It is when we are embodied, and become subject to the needs and inclinations that come with having a body, that there are obstacles that allow the strength of our commitment to the moral law to be measured. For now the commands of reason are forced to compete with the claims of happiness: “Now reason issues its precepts unremittingly, without thereby promising anything to the inclinations, and so, as it were, with disregard and contempt for these claims, which are so impetuous and besides so apparently equitable (and refuse to be neutralized by any command)” (4:405). And now the precise nature of our relationship with the moral law begins to matter. The finite holy will would be an example of a human being who has adopted a fundamental maxim that is morally good. As such, he is resolute in his commitment to the moral law. It matters little to the finite holy will that he is embodied, even though he is subject to the needs and inclinations that come with having a body like any other human being, because these needs and
inclinations can never tempt him to violate duty (6:383). By contrast, a human being who has adopted a fundamental maxim that is morally evil has a relationship with the moral law that is tenuous and easily shaken. When disembodied, his relation to the moral law is never put to the test, and he follows the commands of reason unreflectively, because they are the only commands available for him to follow. But once he is embodied, he becomes subject to the needs and inclinations that come with having a body. He has now been provided with the opportunity to go wrong, because his embodiment leaves him with needs and inclinations that tempt him to do otherwise than his duty, and his capacity to resist them is only as strong as his relation to the moral law.\(^{59}\) It is in becoming embodied that the natural dialectic is generated.

This natural dialectic that arises as a result of our embodiment makes possible the frailty of human nature: “… I incorporate the good (the law) into the maxim of my power of choice; but this good, which is an irresistible incentive objectively or ideally (in \textit{thesis}), is subjectively (\textit{in hypothesis}) the weaker whenever the maxim is to be followed” (6:29). Frailty, as the general weakness of the human heart in complying with adopted maxims (6:29), is a state of the human being. In particular, frailty is a state of moral conflict, in which the sensuous incentives the human

\(^{59}\) To be precise, the human being’s capacity to resist the inclinations that he has depends on his commitment to either the moral law or to prudence. In the Collins lecture notes on ethics, he writes Kant as saying: “There is in man a certain rabble element which must be subject to control, and which a vigilant government must keep under regulation, and where there must even be force to compel this rabble under the rule in accordance with ordinance and regulation. This rabble in man comprises the actions of sensibility. They do not conform to the role of the understanding, but are good only insofar as they do so” (27:360). This self-control is a matter of having discipline, which is “the executive authority of reason’s prescription over the actions that proceed from sensibility” (27:360). The human being that has discipline disciplines himself according to either the rules of prudence (pragmatic discipline) or the rules of morality (moral discipline). So it is written: “Hence we say that self-command consists in this, that we are able to subject all principia to the power of our free choice. This may be considered under two rules, namely those of prudence and morality. All prudence rests, indeed, on the rule of the understanding; but in the rule of prudence understanding is the servant of sensibility, providing it with means whereby the inclination is satisfied, since in regard to ends it is dependent on sensibility. But the true self-mastery is moral in character. This is sovereign, and its laws hold a categorical sway over sensibility, and not as the pragmatic laws do, for there the understanding plays off one sensible factor against the rest. But in order for it to have a sovereign authority over us, we must give morality the supreme power over ourselves, so that it rules over our sensibility” (27:360-361). When morality has supreme power over the human being, it can be said that there is an autocracy of practical reason (6:383).
being experiences threatens to undermine his resolution to do good. The state of frailty precedes
both the choice the human being makes to stand firm against the sensuous incentives that tempt
him astray, and the choice the human being makes when he succumbs to those same sensuous
incentives. For Kant, the moral conflict that is described here is the same, whether it is called frailty
or it is described as a natural dialectic between happiness and morality. And the origins of this
moral conflict lie in the fragility of the human being’s commitment to the good (the moral law),
because had the human being been resolute in his commitment to the moral law in the formation
of his fundamental maxim, then his embodiment and its attendant introduction of sensuous
incentives would not have given rise to this internal conflict. The irresistible incentive objectively
or ideally (in thesi), would have remained subjectively (in hypothesi) the stronger whenever the
maxim is to be followed. But because he had been irresolute in his commitment to the good (the
moral law) in the formation of his fundamental maxim, his embodiment and its attendant
introduction of sensuous incentives revealed this weakness that was already present in his heart.
And so the irresistible incentive objectively or ideally (in thesi), becomes subjectively (in hypothesi)
the weaker whenever the maxim is to be followed. Following Michalson, we might say
that the Willkür’s choice of how it relates to the moral law determines whether or not a moral
conflict will arise when it becomes embodied. And for the human being who has to struggle to
obey the commands of reason against the temptation of sensuous incentives (and who perhaps fails
to entirely and succumbs to the temptation), the body merely provides the occasion for this conflict
through its introduction of incentives that are not in themselves moral, even though they can be
used for moral purposes.60

60 The human being is embodied into an animal body. Because of this, he has a predisposition to animality. And
the predisposition to animality comprises of the instincts “first, for self-preservation; second, for the propagation of
the species, through the sexual drive, and for the preservation of the offspring thereby begotten through breeding;
third, for community with other human beings, i.e. the social drive” (6:26). If the human being does not wholeheartedly
Depravity and Self-Conceit

There are distinctions, Wood observes, in the way Kant uses the term “love” in his practical philosophy (Wood 1996, p. 145). For Kant, love can mean benevolence (Wohlwollen or benevolentia) or well-pleasedness (Wohlgefallen or complacentia). Benevolence has to do with conduct. It is an action that we take to benefit another, as Kant writes: “… Unselfish benevolence toward human beings is often (though very inappropriately) also called love; people even speak of love which is also a duty for us when it is not a question of another’s happiness but of the complete and free surrender of all one’s ends to the ends of another (even a supernatural) being” (6:401). Well-pleasedness, by contrast, is a feeling that we have towards another: “… Only the love that is delight (amor complacentiae) is direct. But to have a duty to this (which is a pleasure joined immediately to the representation of an object’s existence), that is, to have constrained to take pleasure in something, is a contradiction” (6:402). These two kinds of love are often connected. For “to love someone, in other words, is to be pleased by them, or by their perfections, and on this basis to have an inclination toward their good” (Wood 1996, p. 145). But benevolence can also be a duty, and so must also be possible to be benevolent to another out of duty, and not because we are well-pleased by them, because, “What is done from constraint, however, is not done from love” (6:401). Wood suggests that Kant uses the term “self-love” in much the same way: “Since both sorts of love, love as pleasure and benevolent love, apply to oneself, ‘self-love’ is also an ambiguous term, denoting a similarly complex phenomenon. Benevolence toward oneself Kant embrace the moral law, he deprives himself of the moral strength necessary to take control of these animal instincts. First, the instinct for self-preservation – this becomes the bestial vice of gluttony when not constrained (6:27). Second, the instinct for the propagation of the species – this becomes the bestial vice of lust when not constrained (6:27). Third, the instinct for community with other human beings – this becomes wild lawlessness (in relation to other human beings) when not constrained (6:27). This is what it means for evil to be grafted onto the predisposition to animality. The bestial drives are not natural, in that they do not of themselves issue from the predisposition to animality as a root, but are grafted onto the predisposition as a result of the human being failing to control his animal nature.
calls ‘love of oneself’ \textit{(Eigenliebe or philautia)}, while well-pleasedness in oneself is called ‘self-conceit’ \textit{(Eigendünkelt or arrogantia)}” \cite{Wood1996,p.145}. From this, we might make a distinction between benevolence to oneself and well-pleasedness with oneself in terms of the former being a matter of conduct and the latter being a feeling.\textsuperscript{61} Now Kant indeed says that the human being has duties to himself \cite{6:417}. And so we can be benevolent to ourselves (self-love) out of duty we have to ourselves (respect), or because we are well-pleased with ourselves (self-conceit).

In his discussion of self-love in the \textit{Religion}, we find that Kant divides self-love into \textit{mechanical} self-love and \textit{comparative} self-love. The former, mechanical self-love, is described in the following manner: “The predisposition to animality may be brought under the general title of physical or merely mechanical self-love, i.e. a love for which reason is not required” \cite{6:26}. Kant therefore connects this self-love to the instinct that can also be found in animals, particularly the animal instincts for self-preservation, procreation, and community \cite{6:26}. The latter, comparative self-love, is described in the following manner: “The predispositions to humanity can be brought under the general title of a self-love which is physical and yet involves comparison (for which reason is required); that is, only in comparison with others does one judge oneself happy or

\textsuperscript{61} Wood interprets the relation between benevolence and well-pleasedness differently. He does not draw a distinction between benevolence out of duty and benevolence out of inclination (well-pleasedness), but rather connects well-pleasedness to benevolence as cause to effect: “Benevolence is an inclination or wish to benefit someone; well-pleasedness is a pleasure taken in the person, chiefly in the person’s perfections. The two sorts of love are distinct, but they appear to be determinately related. The love of well-pleasedness is the basis for the love of benevolence. To love someone, in other words, is to be pleased by them, or by their perfections, and on this basis have an inclination toward their good” \cite[p. 145]{Wood1998}. Although it is certainly true that the love of well-pleasedness is oftentimes the basis for the love of benevolence, it cannot be the \textit{sole} basis for the love benevolence. This is the entire point of Kant’s discussion of love of human beings – that duty can form the basis for the love of benevolence the same way the love of well-pleasedness can. Hence: “So the saying “you ought to \textit{love} your neighbors as yourself” does not mean that you ought immediately (\textit{first}) to love him and (\textit{afterwards}) by means of this love \textit{do good} to him. It means, rather, \textit{do good} to your fellow human beings, and your beneficence will \textit{produce} love of them in you (as an aptitude of the inclination to beneficence in general)” \cite[6:42]{6:417}. Even though, as Wood notes, benevolent conduct is not love in the proper sense of the word, it is benevolent conduct and not the feeling of delight that is the primary sense of “\textit{love}” in Kant’s interpretation of the command “you ought to \textit{love} your neighbor as yourself”, and the basis of this love is duty and not well-pleasedness in the perfections of our neighbor.
unhappy” (6:27). Notice that the concept of happiness that arises from the predisposition to humanity has a comparative component added to it. It can no longer be construed in purely hedonistic terms as “the satisfaction of all our inclinations (extensive, with regard to their manifoldness, as well as intensive, with regard to degree, and also protensive, with regard to duration)” (A806/B834). As Wood points out, this comparative component does not merely involve the comparison of the different objects of our inclinations, in order to figure out how to best satisfy the system of inclinations that Kant calls happiness (Wood 1996, p. 146). It is in comparison with others that we judge ourselves to be happy or unhappy: “Our pursuit of happiness thus also always involves comparing ourselves with other people; further, the original point of considering ourselves to be happy is that we want to think of ourselves as better than others; conversely, Kant says, we think of ourselves as unhappy or badly off to the extent that we think of our condition as one that might cause others to despise us” (Wood 1996, p. 145).

Self-conceit is intimately connected to the comparative self-love that arises from the predisposition to humanity. As noted by Wood, “Love of oneself is not merely an animal impulse directed toward certain results. Rather, happiness is an end I set for myself, taking it to be objectively good; accordingly, self-conceit not only is a desire to be worth more than others, but also makes an objective claim to superior self-worth” (Wood 1996, p. 147). The human being naturally has inclinations that need to be satisfied. The predisposition to humanity is accompanied by a comparative aspect that impels the human being to compare himself to others in order to determine whether or not he is happy. And he compares himself to others based on the satisfaction of his inclinations – with regard to their manifoldness, with regard to their degree, and with regard to their duration (A806/B834). The human being regards himself as happy, when in comparison to his neighbor, more of his inclinations are satisfied (extensive), when his inclinations are better
satisfied (intensive), and when his inclinations are satisfied longer (protensive). And he regards himself as unhappy to the extent these are not the case. In comparing himself to others on the basis of the satisfaction of his inclinations, the human being comes to adopt a system of valuation based on these inclinations. According to Wood, the relationship between benevolence and well-pleasedness consists in the fact that we are well-pleased in ourselves, and use that as the basis for being benevolent to ourselves. But this well-pleasedness with ourselves would not be possible if there was nothing in us to be objectively well-pleased with, in contradistinction to those around us. This is why the development of self-conceit is accompanied by the formation of a system of values in which it is possible for some human beings to be worth more than others, as Wood writes: “Thus, if (as Kant thinks) our desire for happiness is naturally a desire to be happier than others, and this desire takes the form of a conviction that our happiness is objectively good, then it must be founded on the conviction that we are worth more than others” (Wood 1996, p. 148).

But this is contrary to Kant’s discussion of worth in the *Groundwork*. The Formula of Humanity argument in which Kant identifies humanity as something having unconditional worth is, by and large, read as an argument involving a regress of conditions, as suggested by Korsgaard. Now the argument begins by considering what might be described as putative candidates for being objects of unconditional worth: objects of inclinations, inclinations, natural beings or “things”, and

---

As Reath writes in “Kant’s Theory of Moral Sensibility”: “People naturally place a special importance on themselves, and often make a concern for others conditional on its congruence with their own interests. As Kant understands it, self-conceit tries to get others to accord the same priority to your interests that you give them, by putting your desires forward as conditions on the satisfaction of their own. Though aimed at increasing one’s welfare, it does so by claiming a certain kind of value for one’s person relative to others. How could you possibly get other rational individuals, with desires of their own, to treat your desires as reasons for their actions? Self-conceit attempts to get others to defer to your interests by ranking yourself higher, and by claiming a special value for your person. In this way, it seeks a kind of respect that moves in one direction” (Reath 1989, p. 293). But if the only system of valuation is based on the intrinsic worth of the individual, then self-conceit can gain no traction, because the humanity in every individual is of absolute worth. It is only through the creation of a system of valuation based on the extrinsic characteristics of individuals that self-conceit is able to introduce (artificial) differences in the worth of distinct individuals, without which it would be impossible to claim “a special value for your person” (Reath 1989, p. 293).
persons. The process has us begin with the objects of our choice, where we find that, in order to provide justification for their goodness (or worth), we need to appeal to the inclinations that pushed us to make the choice in the first place. That is, it is because we have a particular inclination that the object of the inclination has any worth whatsoever, because had we not had the inclination, we would have no reason to assign worth to the object at all (Korsgaard 1996, p. 121). But the inclinations themselves cannot be the source of worth either, because Kant observes, “... The inclinations themselves, as sources of needs, however, are so lacking in absolute worth that the universal wish of every rational being must be indeed to free himself completely of them” (4:428). Korsgaard refrains from outright agreement with Kant here. Instead, she brings up bad habits as examples of inclinations that we might wish to be completely free of: “Take the case of a bad habit associated with a habitual craving – it would not be right to say that the object craved was good simply because of the existence of the craving when the craving itself is one you would rather be rid of” (Korsgaard 1996, p. 121). Rather than discuss natural beings, however, Korsgaard moves to the step in which Kant establishes rational nature as the objective end: “[Rational natures], therefore, are not merely subjective ends, the existence of which as an effect of our action has a worth for us, but rather objective ends, that is, beings the existence of which is in itself an end and indeed one such that no other end, to which they would serve merely as means, can be put in its place, since without it nothing of absolute worth would be found anywhere...” (4:428). In other words, Kant finds something in rational nature that differentiates it from everything else in nature as an objective end and it is this that marks it as something that is not to be used merely as means. But what is it about rational nature that can accord it such a privileged status in the system of ends? The way Korsgaard reads it, it is because every agent acts as if his rational nature actually does confer worth upon the objects of its choice (Korsgaard 1996, p. 123). But rational nature being
what it is, there is no relevant difference between distinct rational natures that would privilege one agent’s rational nature over the rational natures of any other (Korsgaard 1996, p. 123). In other words, rational consistency demands that, if an agent considers his own rational nature to be the sufficient condition of worth of the object of its choice, he must consider all other rational natures to be the sufficient conditions for the worth of the objects of their choice. Thus, Korsgaard writes: “If you view yourself as having a value-conferring status in virtue of your power of rational choice, you must view anyone else who has the power of rational choice as having, by virtue of that power, a value-conferring status” (Korsgaard 1996, p. 123).63

Self-conceit is more than simply having an overly high opinion of oneself. It represents an alternative system of values that is diametrically opposed to the system of values expressed in the Formula of Humanity. This the point Wood is making in the following passage:

People are endowed unequally by both nature and fortune, allotted different degrees of strength, beauty, intelligence, talents of all kinds, varying temperaments, different shares of happiness. In Kant’s view, however, none of this has anything to do with their true self-worth, which is always equal. Yet the worth of different rational beings is not merely equal – as if there were some privileged respect in which we could compare their worth and find it present in equal measure. Rather, the dignity of rational nature implies that every rational being’s worth is absolute, beyond the reach of any possible comparative measure. Kant’s view is that all the normal comparative measures of self-worth that people use – honor, wealth, power,

63 Wood disagrees with Korsgaard on this point. For Korsgaard, it is the presence of the predisposition to humanity that marks a human being as an end in itself, and not merely as a means to be used by this or that will at its discretion. But the rational nature that is signified by the predisposition to humanity is to be distinguished from the moral nature that is signified by the predisposition to personality, as Kant explains in the following footnote: “For from the fact that a being has reason does not at all follow that, simply by virtue of representing its maxims as suited to universal legislation, this reason contains a faculty of determining the power of choice unconditionally, and hence to be “practical” on its own; at least, not so far as we can see. The most rational being of this world might still need certain incentives, coming to him from the objects of inclination, to determine his power of choice” (6:26n). It is the predisposition to personality that allows respect for the moral law to be of itself a sufficient incentive to the power of choice for the human being (6:27). And it is the presence of the predisposition to personality, not the predisposition to humanity, that marks a human being as an end in itself for Wood. He writes: “... Kant seems to be saying that dignity belongs to rational beings not in virtue of their end-setting capacity, but in virtue of their capacity to be moral agents and actualize the good will: “Thus morality, and humanity, so far as it is capable of morality, alone have dignity” (G 4:435). The end in itself, he says, “can never be other than the subject of all possible ends themselves, because this is at the same time the subject of a possible will that is absolutely good” (G 4:437). These passages suggest that it is not so much humanity (the capacity to set ends) that has dignity as personality (the capacity to act from pure reason alone and have – or be – a good will)” (Wood 1998, p. 149).
social status in all its forms (success, prestige, attractiveness, charm, charisma, even success in relationships with others) – are expressions of an utterly false sense of values (Wood 1996, pp. 149-150).

From this, we can see that there are two ways in which self-conceit is opposed to the system of values of the moral law. First, self-conceit judges worth on an empirical basis. The things that have worth for it are empirical things like strength, beauty, intelligence, talents of all kinds, varying temperaments, and different shares of happiness (Wood 1996, p. 149). From the standpoint of the moral law, the only thing that has worth is the good will, and this worth is determined independently of all empirical conditions: “A good will is not good because of what it effects or accomplishes, because of its fitness to attain some proposed end, but only because of its volition, that is, it is good in itself; regarded for itself, is to be valued incomparably higher than all that could merely be brought about by it in favor of some inclination and indeed, if you will, of the sum of all inclinations” (4:394). Second, self-conceit judges all worth to be commensurable. This is to judge the worth of all human beings as something that can be compared, as opposed to judging the worth of each human being as beyond the reach of any possible comparative measure. The

---

Mariña has described this opposition between self-conceit and the moral law in terms of the self-transcendence of the will. When the will is autonomous, it finds the law of its determination in the moral law, and it is by obeying the commands of this moral law that a will is able to surpass its finitude and particularity: “For Kant the locus of self-transcendence is not speculative but practical. It is grounded in the unconditional demand that the moral law places upon us. The nature of this unconditional demand is such that it immediately places the self in relation to other rational selves. This relation is, moreover, not one in which the other is valued in terms of the self’s preexisting projects and concerns. Rather, the other must be valued as an end in him or herself” (Mariña 2001, p. 181). By contrast, a will that is heteronomous finds the law of its determination in the inclinations, that is, in the subjective desires that the individual happens to have at the moment of acting. In so doing, “the other is valued in terms of the self’s preexisting projects and concerns” (Mariña 2001, p. 181). As Mariña explains it, this is precisely not to surpass the finitude and particularity of an individual will. “… The decision to attribute value only to the desires one happens to have has a determinate influence on the content of those desires. This influence is such that in cases wherein action is simply in accordance with the strongest desire one happens to have, the content of one’s valuing never moves past the circle of one’s self-concern” (Mariña 2001, p. 185). It is by acting on the basis of objective principles that the will is able to transcend itself. Self-conceit, as the will’s attempt to make its subjective principles the ground of absolute value, is an obstacle to the self-transcendence of the will: “… Self-conceit is the attempt to make the conditioned desires of the finite self absolute, that is, to give them a universal validity that they cannot, in principle, have. Insofar as its principle dynamic is that of decreeing “subjective conditions of self-love as laws”, it rests on a lie in the attempt to make that which is only subjective into something objective. As such, this attitude blocks self-transcendence, that is, the move beyond the finite self and its concerns toward an ability to value that which can in principle be shared by others as a standard of value” (Mariña 2001, pp. 190-191).
point Wood is making here is that, not only is it incorrect to accord human beings more or less worth based on empirical conditions, it is also not correct to treat human beings as if they all have an *equal amount* of worth. For this implies that there is still a way to measure worth when we make a moral decision, the way a utilitarian might choose to act for the benefit of the many, even if it might infringe on the dignity of the few, as if we could tally up the moral worth of those affected by our decision, and measure them against each other. This would still represent the comparative mindset that the moral law opposes. The worth of a human being, according to the moral law, is absolute, and this means that the many are no more worthy of beneficence than the few, despite their greater numbers.

This self-conceit is the evil of depravity, which Kant describes in the *Religion* in the following passage: “... The depravity (*vitiositas, pravitas*) or, if one prefers, the *corruption* (*corruption*) of the human heart is the propensity of the power of choice to maxims that subordinate the incentives of the moral law to others (not moral ones)” (6:30). Insofar as self-conceit is a

---

Wood represents the place of self-conceit in Kant’s moral psychology different. He identifies self-conceit with *radical evil* and not with *depravity*, as we have done here. According to Wood, they are one and the same thing: “We have seen that self-conceit is a propensity to regard our own inclinations as legislative, thereby placing our comparative self-worth above the absolute worth other rational beings have according to the moral law. This very same propensity to invert the rational order of incentives, to place the incentives of our inclination ahead of those of the moral law, is what Kant identifies in the *Religion* as the innate propensity to radical evil in human nature, which we have as rational-social beings who compare ourselves to others and seek superiority over them” (Wood 1998, pp. 153-154). This is because Wood interprets radical evil as an anthropological thesis. Radical evil arises as the result of the social condition of human beings. The solitary human being is moderate in his desires and disposed to contentment (Wood 1999, p. 288). It is when the human being lives alongside other human beings that his desires go out of control and he is thrown into a state of discontent. For he develops an idea of happiness that is *comparative* amidst society, and judges himself happy or miserable in comparison to other human beings. This comparative standpoint comes into conflict with the moral standpoint, as Wood describes in the following passage: “The thesis of radical evil represents human nature as doomed to a basic internal conflict between two equally necessary, but utterly irreconcilable standpoints regarding our own self-worth. It is nature’s plan that our reason should develop through antagonism, that we should measure our worth by comparing ourselves with others and seeking superiority over them. But as reason does develop, it makes us aware that our true self-worth is of an entirely different nature, absolute and incomparable, and moreover that this very same incomparable worth belongs to all those fellow beings our natural-social self-love causes us to see as rivals and obstacles to our self-worth” (Wood 1998, p. 154). Fundamentally, Wood does not think of radical evil in terms of a noumenal choice of an evil fundamental maxim. Thus, he identifies radical evil with self-conceit, as opposed to identifying radical evil as the reason self-love turns into self-conceit, or as the reason the natural inclination to gain worth in the opinion to others gives way to an unjust desire to acquire superiority for oneself over others (6:27).
propensity to regard one’s own inclinations as legislative, Wood describes self-conceit as “placing our comparative self-worth above the absolute worth other rational beings have according to the moral law” (Wood 1996, p. 153). And this is to describe depravity’s inversion of the ethical order of incentives from the perspective of the Formula of Humanity. As an alternative system of values, self-conceit provides the human being with the subjective ground that makes it possible for him to subordinate the incentives of the moral law to others (not moral ones). That is, self-conceit is a psychological framework under which the human being is able to think of himself as having more moral worth than others, and on the basis of which he elevates self-love into a lawgiving and unconditional practical principle (5:74). But at the same time, self-conceit is an outgrowth of the comparative apparatus that comes with the human being’s predisposition to humanity. The concept of happiness that is comparative has to enter the picture before self-conceit does, because self-conceit involves an overstepping of (moral) boundaries: “Out of this self-love originates the inclination to gain worth in the opinion of others, originally, of course, merely equal worth: not allowing anyone superiority over oneself, bound up with the constant anxiety that others might be striving for ascendency; but from this arises gradually an unjust desire to acquire superiority for oneself over others” (6:27). Every human being has a right to pursue happiness (self-love), but

---

66 Depravity is the propensity of the power of choice to maxims that subordinate the incentives of the moral law to nonmoral incentives (6:30). It arises from the predisposition to humanity. The predisposition to humanity gives rise to an inclination to gain worth in the opinion of others (6:27). This inclination to be on par with other human beings, and not allow anyone superiority over oneself, gives way to an unjust desire to acquire superiority for oneself over others (6:27). But this unjust desire is not natural to the predisposition to humanity. Like the bestial vices, it does not issue from the predisposition as a root (6:27). The predisposition to humanity merely instills in us an inclination to gain worth in the opinion of others – the unjust desire to acquire superiority for oneself over others is then grafted onto this inclination (6:27). If the human being does not wholeheartedly embrace the moral law, he deprives himself of the moral strength necessary to go beyond the (moral) limits set by the inclination, from the simple desire to not allow anyone superiority over oneself to an unjust desire to acquire superiority for oneself over others. This is how vices come to be grafted onto inclinations that are originally good. Ambition is not the root of evil that Wood says it is (Wood 1999, p. 290). It is good, because without ambition, the progressive development of the human race would never take place. But ambition unconstrained by reason or morality is either envy (Eifersucht) or haughtiness (Hochmuth, superbia), and these are vices that are grafted upon the inclination as vices of culture (cf. Wood 1999, p. 264).
not to pursue happiness at the expense of others (self-conceit). This is why pure practical reason merely *infringes upon* self-love (5:73). For the human being is allowed to be benevolent towards himself, so long as this does not result in *depravity* towards his fellow man. But pure practical reason *strikes down* self-conceit altogether (5:73). Because self-conceit represents a rival system of values that permits and legitimizes depravity against one’s fellow man.

**Conclusion**

The human being must decide for himself his relationship to the moral law. He can accept the moral law into his heart completely, wholeheartedly. And once he has committed himself to the moral law in this manner, his commitment will be unwavering. This is what it means to have moral strength. Should he accept the moral law into his heart halfheartedly, however, then he has made no real commitment to the moral law. That moral strength is absent here. There is only the logical opposite of moral strength, the absence of moral strength (moral weakness). But strength can only be measured when there is an obstacle that it has to overcome. These obstacles are the real opposites of moral strength. The human being encounters two obstacles that test his moral strength. There are the *sensuous impulses* that he experiences as the result of being embodied. These impulses represent the claims of happiness (4:405). And they act as a powerful counterweight to the demands of duty (4:405). Embodiment – the condition of having a body – belongs with necessity to the possibility of being human. Human beings are not disembodied by nature. But with embodiment comes sensuous impulses and an innate desire for happiness. If the human being is not firmly committed to the moral law, these impulses will take root in the human heart alongside it, insofar as the human being that is morally weak with respect to sensuous impulses is *frail*. There is also *self-love*, which arises from the human being’s original predisposition to humanity. It is natural for a human being to love himself, and to have an interest
in his own happiness. But the human being’s pursuit of happiness has to be restrained, because it cannot be allowed to infringe on the dignity of other human beings, lest self-love turn into self-conceit. If the human being is not firmly committed to the moral law, he will not have the self-restraint to prevent his self-love from becoming self-conceit, insofar as the human being that is morally weak with respect to self-love is depraved. Frailty and depravity are therefore not so different after all. For they are both rooted in a moral weakness that arises from the attitude we have adopted with respect to the moral law, which we have made into a universal rule for ourselves, according to which we will to conduct ourselves, an evil fundamental maxim.67

---

67 For the purposes of continuity, we have focused our attention on frailty and depravity. This means that there has been no discussion at all of impurity, which Kant defined so: “… The impurity (impuritas, improbitas) of the human heart consists in this, that although the maxim is good with respect to its object (the intended compliance with the law) and perhaps even powerful enough in practice, it is not purely moral, i.e. it has not, as it should be, adopted the law alone as its sufficient incentive but, on the contrary, often (and perhaps always) needs still other incentives besides it in order to determine the power of choice for what duty requires; in other words, actions conforming to duty are not done purely from duty” (6:30). In this passage, Kant draws a distinction between actions that are done merely in conformity with duty and actions that are done from duty. Both of these actions are lawful but only the actions that are done from duty are moral, per the Metaphysics of Morals: “The mere conformity or nonconformity of an action with law, irrespective of the incentive to it, is called its legality (lawfulness); but that conformity in which the idea of duty arising from the law is also the incentive to the action is called its morality” (6:219). The lawfulness of actions falls under the doctrine of right. It is therefore concerned with “the external and indeed practical relation of one person to another, insofar as their actions, as deeds, can have (direct or indirect) influence on each other” (6:230). Whereas the morality of actions falls under the doctrine of virtue. It is concerned with the incentives of our actions: “The doctrine of right dealt only with the formal condition of outer freedom (the consistency of outer freedom with itself if its maxim were made universal law), that is, with right. But ethics goes beyond this and provides a matter (an object of free choice), an end of pure reason which it represents as an end that is also objectively necessary, that is, an end that, as far as human beings are concerned, it is a duty to have” (6:380). For this, virtue is needed. Remember that virtue is the capacity and considered resolve to withstand what opposes the moral disposition within us (6:380). And the opponents to the moral disposition of the human being are the sensible inclinations: “… For since the sensible inclinations of human beings tempt them to ends (the matter of choice) that can be contrary to duty, lawgiving reason can in turn check their influence only by a moral end set up against the ends of inclination, an end that must therefore be given a priori, independently of inclinations” (6:380-381). If the human being does not make the noumenal choice to wholeheartedly embrace the moral law, he would not have the capacity and considered resolve to withstand the sensible inclinations that oppose the moral disposition in him. Should this happen, he will be tempted to ends that are contrary to duty, and even though the actions he undertakes as means to this end may still be constrained, either by others or by his innate sense of duty, the action can always only be legal, and not moral. And this is precisely what defines the evil of impurity.
CHAPTER 4. INTERPRETATIONS OF RADICAL EVIL

Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could.

Robert Frost, The Road Not Taken.

Introduction

A survey of the literature on Kant’s doctrine of radical evil will reveal their shared interest in the following question: What is the missing formal proof of the radical evil in human nature in the Religion? The numerous attempts to supply this missing formal proof invariably use the same strategy: they search for a characteristic that is universally present in human beings, and then they argue that simply having this characteristic makes the human being evil. For Firestone and Jacobs, this characteristic is every human being’s participation in a secondary substance, humanity. Whereas Morgan considers this characteristic to be the spontaneous will every human being has as a free agent. Allison tells us that this condition lies in the finitude characteristic of every human being. And Wood traces radical evil to the social condition in which every human being finds himself. The problem is that, because this characteristic is universally present in human beings, our possession of it is not up to us. And this leads us to conclude that our radical evil does not involve the use or abuse of our power of choice in any way whatsoever. In Firestone and Jacobs, the Willkür of humanity chooses evil, but individual human beings are evil simply by being a member of the human species, not because they choose evil for themselves. Morgan wants to identify the propensity to evil with the desire for unrestricted freedom that every free will has by virtue of its spontaneity, but the mere possession of something cannot be culpable when it is given to us by nature and not brought upon ourselves by our choices. The finitude of the human will,
according to Allison, brings with it certain subjective hindrances. But this finitude is analytically contained in the concept of a human being, and so holiness becomes impossible for us by definition. And Wood’s idea that the propensity to evil arises in the social context gives it an origin in time, and so it is something that happens, with a previous state upon which it follows without exception according to a rule (even if we cannot specify what this previous state is). All these attempts to supply the missing proof of the radical evil in human nature try to prove that human nature is radically evil by proving the necessity and universality of the evil in human beings. In so doing, they invariably turn evil into something that happens to us, rather than something that we freely choose.

Chris Firestone and Nathan Jacobs

Firestone and Jacobs, in their book In Defense of Kant’s Religion, interpret the propensity to evil in terms of its relation to the human species. By this they mean that the human species can be considered a subject in its own right, and on this basis they distinguish between the individual

---

Mariña has further argued that our interpretation of the doctrine of radical evil is also guided by our (tacit) acceptance of the Augustinian model of understanding human evil and how it fits in with human destiny. The Augustinian model is described as having the following characteristics: “For our purposes, what is important about this model is the inexplicability and seriousness of the fall to evil. First, the fall is inexplicable. If these beings are fully mature, then presumably this implies they have an understanding of what is of true value. Insofar as they are in the presence of God, they are enjoying the beatific vision. But something tempts them away from that. How can this be? Second, the fall is serious. Given the original righteousness of these beings, who must have had a deep understanding of the value of that which they eschewed in turning away from the good, the fall is an extremely serious moral evil, so serious, in fact, that in some minds it justified the eternal damnation of fallen humanity” (Mariña 2017, p. 185). By contrast, the Irenaeian model of understanding human evil and how it fits in with human destiny can be described in the following way: “God does not create creatures fully formed, but rather as simple and immature. Virtue and knowledge, the perfection of the creature, these are things the creature must attain for itself, that is, it must make itself into what is intended to be through its own development of the seeds of goodness implanted within it. Without such self-development, there is no genuine knowledge or understanding of what has true worth cannot be truly interiorized” (Mariña 2017, pp. 185-186). The inexplicability of radical evil is not a feature of Kant’s doctrine of radical evil, but of the Augustinian lens through which we have been wont to interpret it. But under the Irenaeian model, the doctrine of radical evil can be made explicable, because human beings are created as simple and immature, and they therefore do not fully grasp the consequences of their choices (Mariña 2017, p. 186). Therefore, when faced with making a moral choice at the beginning of their development, their immaturity shows itself, and they choose to side with the principle of self-love over the moral law, thereby precipitating their fall to evil (Mariña 2017, p. 187).
and the species in much the same way Aristotle distinguishes between primary substance and secondary substance. For Aristotle, substances are the subjects of predication. The individual man is a subject of predication, for example, as the subject of paleness. But so too is the species this individual man belongs to, insofar as the human species is the subject of rationality. Because the human species can be predicated of the individual man, but the individual man cannot be predicated of anything else, Aristotle calls the individual man a primary substance, and the species he belongs to a secondary substance. Hence: “Substance, in the truest and primary and most definite sense of the word, is that which is neither predicable of a subject nor present in a subject; for instance, the individual man or horse. But in a secondary sense those things are called substances within which, as species, the primary substances are included; also those which, as genera, include the species” (Categories 2a11-16). Firestone and Jacobs consider the propensity to evil to be something that is predicated of the secondary substance, rather than of the primary substance. And because humanity is predicated to every individual man as a member of the species, it can be said that the propensity to evil is universal to every individual man.

Where the interpretation of Firestone and Jacobs treads new ground is in giving the secondary substance agency in producing this propensity. The choice secondary substance makes to adopt an evil supreme maxim is meant to be taken literally, not metaphorically. Firestone and Jacobs are clear on this in the following passage: “… The exercise of Willkür in the context of the supreme maxim is very much like an Adamic Fall into sin. However, rather than the disposition being the product of one particular individual, such as Adam, the dispositional bent is chosen by the secondary substance, human, when first individuated and actualized in matter, prior to all particular exercises of freedom” (Firestone-Jacobs 2008, pp. 147-148). In other words, it is not the individual man that makes the free choice that brings about the propensity to evil in him, but rather
the human species that makes the free choice that brings about the propensity to evil in the species, and the individual man is culpable for this propensity by the mere fact of his participation or membership in the human species. This can be seen in the following passage, where they write: “In this way, our culpability for the disposition resides in our participation or membership in the species. We cannot separate ourselves from our humanity; the properties of our species are our species. This freely chosen disposition is thus part of the species in which we participate and is that which defines our common moral nature” (Firestone-Jacobs 2008, p. 148).

The reasoning Firestone and Jacobs provide for this is relatively straightforward. Kant indicates in the Religion that the human being is by nature evil (6:42). And what it means to be evil by nature is that being evil applies to every human being considered in his species (6:32). But a propensity can either be physical or moral (6:31). If a propensity is physical, then it pertains to a human being’s power of choice as a natural being, and this means that the choice is based on

---

A natural objection to raise at this point would be this: Why should we think that Kant’s doctrine of radical evil depends on an Aristotelian metaphysics of substance? For Firestone and Jacobs want to argue that it is not through the Willkür belonging to the individual (primary substance), but by the Willkür belonging to humanity as a species (secondary substance), that “determines the moral bent of the species prior to all particular exercises of freedom” (Firestone-Jacobs 2008, p. 147). The burden of proof is on Firestone and Jacobs to show that Kant would accept the Aristotelian distinction between primary substance and secondary substance into his metaphysics. They attempt to do so by arguing that the prevailing notion of species during the period was in line with the traditional Aristotelian notion of species:

... Leroy E. Loemker has argued that the dominant concept of species in the early sixteenth century was Aristotelian in nature. Loemker points to Julius Caesar Scaliger as a prime example of the type of understanding dominant in the period: forms are immanent and individuated, being placed as “seeds” within matter, first in potential and brought forth in actu. These “seeds” are then passed on and further individuated in the process of generation. Human concepts are therefore held to be signs of things known through contact with the individuated form. Scaliger’s not uncommon use of Aristotle thus “made Aristotle the founder of the preformation theory of generation”. Loemker notes that this Aristotelian understanding of species persisted into the seventeenth century (albeit with Platonic additions [...]), and he suggest that this understanding of species was dominant in figures such as Leibniz. Moreover, part of this realist view of species was the notion that “ideas are active or, as Leibniz put it, have an exigency to actualize themselves, and are therefore discoverable in things by scientific analysis”. In light of these historical trends preceding Kant, we should not think it strange to find Kant embracing a realist view of species, and understanding immanent form as active (Firestone-Jacobs 2008, p. 147)
sensory inducements (6:31). If a propensity is moral, then it pertains to a human being’s power of choice as a moral being, and this means that the choice presupposes a Willkür that has the capacity to freely make the choice (6:31). Now the propensity to evil is a moral propensity, and this means that it pertains to a human being’s power of choice as a moral being. For the propensity to evil to be part of human nature, however, means that the propensity to evil is not the property of the individual man, but of the human species, and this moral propensity entails that the human species has a Willkür to freely make the choice.

Hence, Firestone and Jacobs write: “Yet, per the concept of nature, this maxim must be an essential property of our species – even if the specific bent of our species is non-essential. Moral nature therefore indicates that the human species, or humanity as secondary substance, must possess Willkür, at least with regard to the single act of generating the supreme maxim that defines our moral nature – a nature for which we are culpable as members of humanity” (Firestone-Jacobs 2008, p. 150).

It is difficult, however, to see how this solution can avoid the incredulous stare. The idea that the human species has its own Willkür that makes the choice of evil in its determination of human nature seems, to most people, metaphysically extravagant at best. To their credit, Firestone and Jacobs appeal to this in order to explain why the propensity to evil cannot be extirpated through human forces: “With Kant’s investigation of the species being outside the individual sphere, the supreme maxim is not an individual affair. If the corrupt maxim was an individual affair, the change of maxim, it would seem, would be within the individual’s power to undo. But since the adoption of the supreme maxim is by the species, it stands beyond the power of the individual to affect” (Firestone-Jacobs 2008, p. 150). This seems to misinterpret the point Kant makes in 6:37. For Kant denies that it is within the individual’s power to undo the corrupted fundamental maxim, even though it is an individual affair. This is because the maxims of a human being are arranged in a hierarchy according to their generality. The more general a maxim is, the higher its position in the hierarchy, and the more numerous the maxims that fall under it (are grounded by it). If there is to be a change to the maxims in the hierarchy, grounds must be given for this change, and this is done by appeal to the maxims higher in the hierarchy that ground it. But the fundamental maxim is located at the apex of the hierarchy. There are no maxims above it in the hierarchy that we can appeal to in order to have it changed. This is why an evil fundamental maxim cannot be extirpated through human forces, as Kant makes evident in the following passage: “Now if a propensity to this [inversion] does lie in human nature, then there is in the human being a natural propensity to evil; and this propensity itself is morally evil, since it must ultimately be sought in a free power of choice, and hence is imputable. This evil is radical, since it corrupts the ground of all maxims; as natural propensity, it is also not to be extirpated through human forces, for this could only happen through good maxims – something that cannot take place if the subjective supreme ground of all maxims is presupposed to be corrupted” (6:37). Nowhere in the passage does Kant go beyond the individual sphere, as Firestone and Jacobs claim he does.

---

70 Firestone and Jacobs appeal to this in order to explain why the propensity to evil cannot be extirpated through human forces: “With Kant’s investigation of the species being outside the individual sphere, the supreme maxim is not an individual affair. If the corrupt maxim was an individual affair, the change of maxim, it would seem, would be within the individual’s power to undo. But since the adoption of the supreme maxim is by the species, it stands beyond the power of the individual to affect” (Firestone-Jacobs 2008, p. 150). This seems to misinterpret the point Kant makes in 6:37. For Kant denies that it is within the individual’s power to undo the corrupted fundamental maxim, even though it is an individual affair. This is because the maxims of a human being are arranged in a hierarchy according to their generality. The more general a maxim is, the higher its position in the hierarchy, and the more numerous the maxims that fall under it (are grounded by it). If there is to be a change to the maxims in the hierarchy, grounds must be given for this change, and this is done by appeal to the maxims higher in the hierarchy that ground it. But the fundamental maxim is located at the apex of the hierarchy. There are no maxims above it in the hierarchy that we can appeal to in order to have it changed. This is why an evil fundamental maxim cannot be extirpated through human forces, as Kant makes evident in the following passage: “Now if a propensity to this [inversion] does lie in human nature, then there is in the human being a natural propensity to evil; and this propensity itself is morally evil, since it must ultimately be sought in a free power of choice, and hence is imputable. This evil is radical, since it corrupts the ground of all maxims; as natural propensity, it is also not to be extirpated through human forces, for this could only happen through good maxims – something that cannot take place if the subjective supreme ground of all maxims is presupposed to be corrupted” (6:37). Nowhere in the passage does Kant go beyond the individual sphere, as Firestone and Jacobs claim he does.
and Jacobs attempt to make the idea more acceptable in the following passage: "A strong history of interpretation exists that takes Aristotle to distinguish an individual (tode ti) from a particular (kath‘hekasta). When a particular is non-repeatable and cannot be predicated of another object, a universal can be individual. If individuation is understood in this way, it is possible to conceive of secondary substance being individuated in the first particular human, actively determining the moral bent of the species prior to all particular exercises of freedom, and then being further individuated in the process of generation, bringing with it an innate disposition" (Firestone-Jacobs 2008, p. 147). The human species, according to this passage, is not an abstraction from individual human beings, but constitutive of individual human beings. The moment humanity first became individuated and actualized in matter, say in the individual known as Adam, that was when humanity used its Willkür to determine the dispositional bent of the species. Through this passage, Firestone and Jacobs hitch their wagons to the organic whole theory of original sin, which has a long and venerable history that includes such proponents as Aquinas and Edwards (Rea 2007, p. 328). The organic whole theory, as described by Rea, is this: “The idea, in short, is that humanity, human nature, or the human race is an organic whole with the following properties: (a) it is a moral agent; (b) every individual human being is a part or instance of it; and (c) it committed the sin of Adam by virtue of having a part or instance – namely, Adam – that committed that sin. On this view, it is by virtue of being parts, instances, or members of this whole that individuals other than Adam participated in Adam’s sin and share the guilt for it” (Rea 2007, pp. 329-330). The idea is that we can make a distinction between Adam’s first sin and Adam’s second sin. Because Adam’s first sin is committed by Adam qua humanity, the sin is imputed to humanity. By contrast, because Adam’s second sin is committed by Adam qua Adam, the sin is imputed to Adam himself.71 But,
as participants and members of the human species, individual human beings other than Adam participated in Adam’s sin and share the guilt for it. Firestone and Jacobs are proposing a Kantian adaptation of the theory. As humanity first became individuated and actualized in matter as the first human being, it had to decide the dispositional bent of the species. The difference between Adam’s first and second exercises of freedom is this: Adam’s first exercise of freedom was made by Adam qua humanity. It is not the product of one particular individual because it was not Adam’s Willkür but humanity’s Willkür that made the decision to choose the propensity to evil. By contrast, Adam’s exercise of freedom was made by Adam qua Adam. It is the product of one particular individual because it was Adam’s Willkür that made the decision, and not humanity’s Willkür.

Thus, as participants and members of the human species, individual human beings other than Adam participated in Adam’s first exercise of freedom and share the guilt for it.72

hard to see why Adam’s first sin, and that sin alone, would involve all of human nature in the way required by the analogy. Even if we grant that there is a sense in which your hand, but not your foot is to blame for sins you commit with your hand, still it is hard to see why Adam’s first sin was a sin committed with his whole nature, as it were, rather than a sin that simply involved him as an individual” (Rea 2007, pp. 330-331). Insofar as Firestone and Jacobs want to claim that the human species was an individual in the person of Adam, because there were no other human beings besides him, it would seem that they cannot give no principled distinction between the Willkür of Adam and the Willkür of the human species. Every exercise of the power of choice by the Willkür of is therefore an exercise of the power of choice by the Willkür of the human species, and vice versa. And there will be no way to separate the Willkür of Adam from the Willkür of the human species, except by separating the substantial being of Adam from that of the human species, which would only happen when the human species becomes instantiated in other individuals besides Adam.

The problem with interpreting the doctrine of radical evil in this way is that moral evil is ascribed to every human being by way of an inheritance from our first parents, Adam and Eve. This is precisely what Kant says radical evil is not. Kant is explicit in his rejection of this idea: “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 quae non fecimus ips, vic ex nostra puto (Race and ancestors, and those things which we did not make ourselves, I scarcely consider as our own)” (6:40). Inherited moral evil, like race and ancestors, are not things that we make or choose for ourselves, and therefore not something we would consider our own. In interpreting the origin of moral evil in terms of an organic whole theory, Firestone and Jacobs consider Adam’s choice to sin to be made through the Willkür of the species (humanity) and not through the Willkür of the individual (Adam), and every human being is guilty of this sin due to his participation of the species. But this has the effect of making moral evil something that is inherited. In particular, the organic whole theory represents the origin of moral evil in human beings as an inherited sin, rather than an inherited disease or an inherited guilt, in accordance with the taxonomy provided by Kant in the footnote: “The three so-called “higher faculties” (in the universities) would explain this transmission each in its own way, namely, either as inherited disease, or inherited guilt, or inherited sin. (1) The Faculty of Medicine would represent the inherited evil somewhat as it represents the tapeworm, concerning which certain natural scientists are actually of the
This solution to the problem of radical evil by Firestone and Jacobs gives rise to the following problem: It seems that I cannot be morally responsible for the decisions made by the Willkür of a secondary substance. They make the following two claims: That it is the human species, and not the individual human being, that determines the dispositional bent of the human species as evil. And that the individual human being is evil because of his participation in the human species. These two claims together imply that the propensity to evil is a physical propensity for the individual human being, not a moral propensity. For remember that good and evil is not imputed to the individual human being unless he has, through his own Willkür, made the choice to incorporate the relevant incentive into his maxim, “… for otherwise the use or abuse of the human being’s power of choice with respect to the moral law could not be imputed to him, nor could the good or evil in him be called “moral” (6:21). But the individual human being has no say in how the Willkür of the human species is exercised, as Rea remarks in his discussion on the organic whole theory: “It is clear even from this rough sketch, the obvious challenge for the view is to explain in what sense, if any, the non-Adamic parts or instances of the whole could have prevented the sin of Adam. Prima facie, they could not have” (Rea 2007, p. 330). Nor does the individual human being have any say in whether or not he wants to be human.\(^73\) And so his participation in

---

\(^73\) It is not strictly correct to say that the human being is blameless for the choice that the Willkür makes on behalf of the entire species. For imagine that a soul can be reincarnated in another body after its death, and the body it is reincarnated into depends on the use or abuse of its power of choice during its previous life, as believed by some Eastern religions. For example, a human being that has lived a moral and upright life might be reincarnated as a deva. Whereas a human being that has lived a life of villainy and depravity might be reincarnated as a hungry ghost. In such a cosmology, a soul is reincarnated as a human being because of the way its power of choice was exercised (used or
the human species, through which he inherits the dispositional bent towards evil, is not up to him either. It is therefore difficult to see how the propensity to evil in an individual human being can in any way be traced to an abuse of his power of choice. And if the propensity to evil does not involve an abuse of the individual human being’s power of choice, then it cannot be imputed to him. And this would mean that the propensity to evil in him cannot be called “moral” – the propensity to evil in the individual human being would have to be a physical propensity and not a moral propensity. That the human species has a moral propensity to evil would be, given the distinction between humanity and its members that has been introduced, completely beside the point.

Seiriol Morgan

The interpretation of the propensity to evil proposed by Morgan in his article “The Missing Formal Proof of Humanity’s Radical Evil in Kant’s Religion” is next up for consideration. The way he explains it, the choice of a fundamental maxim is ultimately a choice between negative and positive freedom, and not a choice between good and evil per se. In the Groundwork, Kant defines negative freedom in terms of freedom from outside determination: “Will is a kind of causality of living beings insofar as they are rational, and freedom would be the property of such causality that it can be efficient independently of alien causes determining it, just as natural necessity is the property of the causality of all nonrational beings to be determined to activity by the influence of alien causes” (4:446). By contrast, positive freedom is defined in terms of the laws that can

abused) in its previous life. From this, it follows that the soul is responsible for the human form it is reincarnated as. Then even though the individual human being had no say in how the Willkür of the human species was exercised before his birth, he can nevertheless be imputed the sin of the species, because he did have a say in his participation in the human species. (If he had acted differently in his previous life, he would not have been reincarnated as a human being, but as some other creature, or perhaps have escaped the cycle of rebirth completely). But it is exceedingly unlikely that a devout Pietist like Kant would have embraced reincarnation as a solution to the universal imputation of the Adamic sin.
properly determine a will that is free in the negative sense: “Since the concept brings with it that of laws in accordance with which, by something that we call a cause, something else, namely an effect, must be posited, so freedom, although it is not a property of the will in accordance with immutable laws but of a special kind; for otherwise a free will would be an absurdity” (4:446). This distinction between negative freedom and positive freedom is important to Morgan’s interpretation of radical evil because of the circumstances in which he thinks the fundamental maxim is chosen. The choice of the fundamental maxim is a choice that precedes every deed that takes place in space and time (6:31). It is therefore a deed that cannot itself take place in space and time (6:31). But having been separated from all spatiotemporal conditions, this choice of a fundamental maxim cannot be made on the basis of our inclinations and desires, as Morgan puts it:

The problem we face is that in imagining the will making a choice between morality and self-love, prior to its immersion in the world, we have pared away virtually every element of the self that might provide it with reasons – its desires, its emotions, its values, its membership in society, its history, its individual identity. All that remains is the sheer power of choice, the will’s spontaneity. And so it is precisely within spontaneity that Kant locates its reasons. Kant’s reasoning is that since all the will is freedom, the only thing that can possibly provide the will with a reason is spontaneity itself. As freedom is its inner nature, the will has reason to choose the principle which best preserves and expresses it (Morgan 2005, p. 77).

According to Morgan, the choice of the will’s fundamental maxim has to be made in terms of freedom. In choosing the moral law, the will chooses freedom in the positive sense as the principle which best preserves and expresses it. But even in the choice of self-love, the will chooses freedom as the principle which best preserves and expresses it, understanding this freedom as freedom in the negative sense, where the will chooses to be free of any and all influences, including the influence of the moral law. As such, the restrictions that the demands of morality places on the freedom of the rational agent comes to be seen as an alien influence, because they stand in the way
of the will’s self-expression, and therefore as a hindrance that is to be removed for the sake of freedom. Thus, the moral incentive that determines the power of choice takes the form of freedom positively conceived, whereas the incentive of self-love that determines the power of choice takes the form of freedom without restraint: “The picture that emerges is of the human being as of her nature inclined to a kind of gratuitous willfulness, in which she simply fetishizes and elevates to a supreme value, trumping all other considerations, the unlimited indulgence of her whims” (Morgan 2005, p. 85).

But how is any of this supposed to show that the propensity to evil is universal? Palmquist’s observation seems apt here: “… He ingeniously presents a definition of outer freedom as self-deceptive (i.e., as excluding the inner freedom of the moral law) and claims (perhaps rightly) that this is an assumption adopted by people who have succumbed to the evil propensity. But how does

---

74 The passage Morgan is referencing here comes from the *Metaphysics of Morals*: “Resistance that counteracts the hindering of an effect promotes this effect and is consistent with it. Now whatever is wrong is a hindrance to freedom in accordance with universal laws. But coercion is a hindrance of resistance to freedom. Therefore, if a certain use of freedom is itself a hindrance to freedom in accordance with universal laws (i.e., wrong), coercion that is opposed to this (as a *hindering of a hindrance to freedom*) is consistent with freedom in accordance with universal laws, that is, it is right. Hence there is connected with right by the principle of contradiction an authorization to coerce someone who infringes upon it” (6:231). Now Kant believes that the principle of right can be analytically derived from the concept of freedom (6:396). From this, it follows that reflection on outer freedom would allow us to discern its limits, that the outer freedom of an individual extends only so far as it does not conflict with the outer freedom of another, beyond which it becomes a hindrance to freedom (6:396). It seems then that the agent who adopts outer freedom must understand that it does not license untrammeled freedom, even when his concept of freedom does not include the concept of inner freedom (a freedom bound by morality). Morgan is aware of this objection, in response to which he writes: “But analysis of the concept merely shows us what rightness is, and nothing about the fact itself compels anyone to take any practical interest in it, and govern his behavior according to the principle of right. Rather, the only thing that could motivate an individual to respect the outer freedom of others with whom he interacts, besides fear of their personal power and that of punitive authority, is morality. From the amoral perspective of a particular agent considering his de facto relationship with other agents, rather than that of someone taking an abstract view of the de jure relations between agents in general, nothing about the idea of outer freedom itself calls for its restriction when it begins to encroach upon the choices of others (Morgan 2005, p. 83). But if the concept of right is analytically contained in the concept of outer freedom, then untrammeled freedom, as something contrary to right, cannot also be part of the concept of outer freedom. Therefore, whatever the noumenal agent is choosing as his incentive in the construction of his fundamental maxim, he is not making his choice outer freedom, because rightness is contained in the concept of outer freedom, and he has no practical in governing his behavior according to the principle of right. Nor is he making the choice of inner freedom, because morality is contained in the concept of inner freedom, and he has no practical interest in governing his behavior according to the principle of morality, per our hypothesis. Therefore, what is the concept of freedom that the will chooses, given that for such a will, “as pure spontaneity is its nature, only freedom could be an incentive for it”, and the concepts of outer freedom and inner freedom are not available to it?
this clarification of the concept amount to a transcendental proof?” (Palmquist 2008, p. 269). To be fair, we find that Morgan actually does link negative freedom to universality in his summary of the argument:

Consequently the argument also establishes the universality of the propensity. We know that at least some wills are attracted to evil, because some agents perform acts that clearly subordinate moral considerations to ones of sensibility. Since sensuous incentives cannot determine the will except insofar as the agent has incorporated them into his maxim, the evil must have been presented with an incentive from its own nature to do so. But its own nature is pure spontaneity, a nature it shares with every other will. Hence if any will is motivated by its own nature as bare freedom to incorporate evil into its maxim, every will must present itself with incentive to do so. And since the incentive emerges from its innermost nature in this way, it does not seem unreasonable to consider it inextirpable, but possible to overcome in practice, again as Kant said (Morgan 2005, p. 87).

Let us try to unpack Morgan’s argument. The propensity to evil is defined here as the incentive for untrammeled freedom. Because some people are actually evil, they must have exercised their free Willkür to subordinate the moral incentive to nonmoral incentives. For this to be possible, they must have adopted an evil fundamental maxim, because it is an evil fundamental maxim that grounds every deed contrary to law (6:31). This choice must itself involve the incorporation of an incentive into the maxim, otherwise it would not be free. But a sensuous incentive would not do here, because there are no sensuous incentives available for incorporation in this intelligible deed. This is why Morgan posits that the spontaneity of freedom must be, by itself, sufficient to generate this nonmoral incentive for the Willkür to incorporate. And this nonmoral incentive is the incentive of untrammeled freedom. Now every free Willkür is spontaneous in the exact same way. It must therefore follow that this incentive of untrammeled freedom is so generated by every free Willkür, leading Morgan to conclude: “Hence if any will is motivated by its own nature as bare freedom to
incorporate evil into its maxim, every will must present itself with incentive to do so” (Morgan 2005, p. 87). And Morgan takes this to show the universality of the propensity to evil.75

But there is something that Morgan does not make clear throughout his entire argument: Are we morally responsible for the propensity to evil in us because we had incorporated the incentive to untrammeled freedom into our maxim or because we simply possess this incentive to untrammeled freedom? Suppose it is the former. Then we are morally responsible for our propensity to evil because we had incorporated the incentive to untrammeled freedom into our fundamental maxim. If so, it would seem that Morgan has gotten the relevant relations reversed: He posits the incentive of untrammeled freedom to explain why we choose self-love when he should have posited the incentive of self-love to explain why we choose untrammeled freedom.76

75 What Morgan writes in its entirety is this: “We know that at least some wills are attracted to evil, because some agents perform acts that clearly subordinate moral considerations to ones of sensibility. Since sensuous incentives cannot determine the will except insofar as the agent has incorporated them into his maxim, the evil will must have been presented with an incentive from its own nature to do so. But its own nature is pure spontaneity, a nature it shares with every other will. Hence if any will is motivated by its own nature as bare freedom to incorporate evil into its maxim, every will must present itself with the incentive to do so” (Morgan 2005, p. 87). Now Morgan is unwilling to say that all wills are attracted to evil (Morgan 2005, p. 64). Therefore he cannot say that every will has subordinated moral considerations to the considerations of sensibility. Instead, he wants to say that, those wills that are attracted to evil, have subordinated moral considerations to the considerations of sensibility, and this is only possible because its own nature as bare freedom has presented it with the incentive to do so. And that every will has this same nature, and every will is therefore presented with the same incentive to subordinate moral considerations to the considerations of sensibility. But this does not establish a propensity to subordinate moral considerations to the considerations of sensibility in every human being, only a possibility of subordinating moral considerations to the considerations of sensibility, adding to the possibility of subordinating the considerations of sensibility to moral considerations, which was already present. A propensity for something to happen makes it more likely to happen – the mere presentation of an incentive is not enough here.

76 According to Morgan, the choice between morality and self-love is made in the noumenal realm, in the form of the agent’s selection of his fundamental maxim:

In effect, the task the will is faced with then is choosing what is going to count as a reason for action for it when it gets into the world. Various reason-generating principles spring to mind, but for the sake of argument, let’s confine ourselves within Kant’s strictures and limit the choice to that of self-love or morality. So the will has to choose what will count as supreme reason for it when it arrives in the world, the satisfaction of its own inclinations or the conformity of its choices with the categorical imperative. In deciding which of the principles to give overriding status, it is deciding which to subordinate to the other. So the choice it has to make is one between a good and evil disposition. But how is it to make the choice? Since it is deciding what is going to count as a reason for it in the future, what reason could it have to guide it now? Certainly no appeal can be made either to morality or self-interest, conceived as the satisfaction of inclination, since any such appeal would be a circular justification. So one might think that the will would be in a condition prior to all reasons, and any choice it makes would simply be a random exercise of its spontaneity, an utterly unmotivated “leap of volition” (Morgan 2005, pp. 76-77 – emphasis mine).
This is because the human being who chooses the incentive of untrammeled freedom does not do so because he mistakenly thinks it is morally good. He does so knowing full well that he has chosen evil, morally speaking. Morgan himself admits this in the following passage: “Motivated by the allure of a distorted picture of freedom, it is blinded to the reality of its choice. But we are certainly not to conceive of the fundamental problem here as a kind of intellectual mistake, as if the evil will were the fruit of a confusion about a matter of fact. The evil person is not someone who has made a subjectively rational decision on the basis of false information, who at most we could convict of culpable ignorance. The evil person is someone who willfully does what in some sense she knows to be bad” (Morgan 2005, pp. 84-85).

Indeed, it cannot be any other way, because if we suppose that the will is free, the law that alone is competent to determine such a will is the law of reason. As Kant writes in the Second Critique:

> Since the matter of a practical law, that is, an object of maxim, can never be given otherwise than empirically whereas a free will, as independent of empirical conditions (i.e., conditions belonging to the sensible world), must nevertheless be determinable, a free will must find a determining ground in the law but independently of the matter of the law. But, besides the matter of the law, nothing further is contained in it than the lawgiving form. The lawgiving form, insofar as this is contained in the maxim, is therefore the only thing that can constitute a determining ground of the will” (5:29).

In the selection of the fundamental maxim, the incentive of morality cannot be used to select the good disposition, and the incentive of self-interest cannot be used to select the bad disposition. Now Morgan does not think that the will, in this state, would be a condition prior to all reasons, and the selection of the good disposition or bad disposition is a matter left to chance. Instead, he argues that the nature of the will as pure spontaneity is itself an incentive that is used in the selection of the fundamental maxim. If the incentive of inner freedom, or a freedom in accordance with the law of reason, is incorporated into the agent’s fundamental maxim as an incentive, then he has a good disposition. But if the incentive of outer freedom, or a freedom that is completely unrestrained, is incorporated into the agent’s fundamental maxim as an incentive, then he has a bad disposition. But given that the notion of inner freedom is that of a freedom in accordance with the law of reason, it seems that morality is already presupposed in the notion itself. Morality is a more fundamental notion than inner freedom or positive freedom. And the same point can be made with respect to self-love and outer freedom, mutatis mutandis.
But the determination of the will by the mere lawgiving form of the law is this: to act only in accordance with that maxim through which you can at the same time will that it become a universal law. This is the moral law, of which we become immediately conscious as soon as we draw up maxims of the will for ourselves (5:30), hence it is given to us first, as the ratio cognoscendi through which we come to know freedom (5:4n). It is therefore impossible for Morgan’s rational agent to be aware of his freedom and yet be ignorant of the content of the moral law. This means that a human being who adopts a freedom that is negative to the exclusion of his positive freedom, does so knowing full well that the conception of freedom he is adopting is incomplete, and that a complete conception of freedom would have to include positive freedom, that is, freedom as autonomy (Morgan 2005, p. 82). In other words, the appeal to untrammeled freedom presupposes that the germ of evil has already taken root in the human heart. The incentive to untrammeled freedom would not be an incentive for us if we had not already elevated our own freedom above the freedom of others. And this is just self-love disguising itself in the false piety of freedom-talk.

But suppose it is the latter. Then Morgan’s position is that the presence of a propensity to evil in us is completely explained by the fact that we possess this incentive to untrammeled freedom. If so, this would seem to make Morgan’s propensity to evil a physical propensity rather than a moral propensity, because we cannot be morally responsible for merely having an incentive that is presented to us from the nature of our wills as pure spontaneity. Remember that we are only morally responsible when we have incorporated the relevant incentives into our maxim through the free exercise of our Willkür. But we are no more responsible for possessing a nonmoral incentive generated from the nature of the will as pure spontaneity than we are responsible for possessing a nonmoral incentive generated from our sensuous nature as embodied creatures. Therefore, if merely having the incentive to untrammeled freedom is sufficient to ascribe to us a
propensity to evil, then Morgan has made the propensity to evil *analytically necessary* for every being with a will that is purely spontaneous (which would presumably include God). For this reason, Morgan falls back on the distinction between freedom in the sense of spontaneity, and freedom in the sense of indifference: “To possess liberty of spontaneity is to be oneself the cause of one’s own choices; liberty of indifference is the power to do otherwise than one actually does, so that one can either φ or not φ” (Morgan 2005, p. 92). It is freedom in the sense of indifference that leads us to think that radical evil being both universal and freely chosen is an exceedingly improbably affair, because “it is antecedently likely that some people would have freely adopted a morally good supreme maxim while others adopted a morally bad supreme maxim” (Quinn 1984, p. 36). But Morgan argues that the two concepts of freedom are not coextensive. That is, the fact that our choice of radical evil was *spontaneously* made does not entail that we ever had the *power to do otherwise*, and it is therefore logically possible for radical evil to be both universal and freely chosen, albeit in the sense of being spontaneously chosen and not in the sense of being chosen with indifference.

To drive home his point, Morgan appeals to God as an example of a being that is free without having the power to do otherwise: “… God for instance is literally incapable of acting contrary to the law, but this certainly does not mean that God is thereby unfree. On the contrary, he is as free as it is possible to be, since there is obviously no question of him betraying the principle through which freedom is actualized” (Morgan 2005, p. 94). But this feels like a red

---

77 The passage Morgan is referencing here is from the *Lectures on the philosophical doctrine of religion*, in which Kant writes: “One might raise the objection that God cannot decide otherwise than he does, and so he does not act freely but from the necessity of his nature. The human being, however, can always decide something else, e.g. a human being, instead of being benevolent in this case, could also not be that. But it is precisely this which is a lack of freedom in the human being, since he does not always act according to his reason; but in God it is not due to the necessity of his nature that he can decide only as he does, but rather it is true freedom in God that he decides only what is in conformity with his highest understanding” (28:1068). Now Morgan is making the point that Kant allows for acts that are free because they are spontaneous, even though we do not have the power to do otherwise with respect to them (Morgan 2005, p. 94). But this seems like a *non sequitur*. In the passage, Kant says that human beings have
herring. For let us concede that God is free, despite not having the power to do otherwise. We may even concede that “Kant maintains a conception of liberty that allows for “originary” acts of freedom, acts that are free because they are spontaneous, even though we do not possess liberty of indifference with respect to them” (Morgan 2005, p. 94). This is a far cry from proving that this conception of liberty that allows for “originary” acts of freedom is applicable to the individual human being’s choice of a fundamental maxim, so that this act would be free because it is spontaneous, even though the human being does not possess liberty of indifference with respect to it. Indeed, Kant says the exact opposite: “Impulses of nature, accordingly, involve obstacles within the human being’s mind to his fulfillment of duty and (sometimes powerful) forces opposing it, which he must judge that he is capable of resisting and conquering by reason not at some time in the future but at once (the moment he thinks of duty): he must judge that he can do what the law tells him unconditionally he ought to do” (6:380). Every human being has an evil fundamental maxim, but it is true for each and every one of them that he should have chosen differently, and because he should have chosen differently, he could have chosen differently. Otherwise there would have been no abuse of the human being’s power of choice, and the propensity to evil would not be something brought by the human being upon himself, but something that happened to the

the power to do otherwise because he does not always act according to reason. This signifies a lack of freedom on the part of the human being. The freedom that the human being lacks is the freedom of the Wille, because only a heteronomous will can depart from reason in this way: “If the will seeks the law that is to determine it anywhere else than in the fitness of its maxims for its own giving of universal law – consequently if, in going beyond itself, it seeks this law in a property of any of its objects – heteronomy always results. The will in this case does not give itself the law; instead the object, by means of its relation to the will, gives the law to it” (4:441). God cannot do otherwise than he does. But Kant considers God to be free because He decides only what is in conformity with His highest understanding. This is the freedom of the Wille, because it is an autonomous will that is subject only to laws given by itself but is still universal (4:432). As Kant explains it: “Autonomy of the will is the property of the will by which it is a law to itself (independently of any property of the objects of volition). The principle of autonomy, therefore: to choose only in such a way that the maxims of your choice are also included as universal law in the same volition” (4:440). The dichotomy that Kant sets up in the passage is between the freedom of autonomy and the lack of freedom in heteronomy, and not between the freedom of spontaneity and the freedom of indifference (the power to do otherwise). While it may well be the case that Kant conceives of freedom in terms of spontaneity (see Sgarbi 2012), the passage that Morgan references from the Lectures on the philosophical doctrine of religion would be of no help establishing any such thing.
human being. And only the former would be a \textit{moral} propensity to evil, the latter being only a \textit{physical} propensity to evil. And this is why Kant writes, when it comes to the \textit{rational} origin of evil in human nature, every derivative sin must be like original sin in that the human being is considered to have had the power to do otherwise (6:44).

\textbf{Henry Allison}

Turning now to Allison’s argument for the universality of the propensity to evil, we find that he begins by contrasting the propensity to evil with the propensity to good. The propensity to good is defined in terms of the possession of a holy will: “A perfectly good will would, therefore, equally stand under objective laws (of the good), but it could not on this account be represented as \textit{necessitated} to action in conformity with law since of itself, by its subjective constitution, it can be determined only through the representation of the good. Hence no imperatives hold for the \textit{divine} will and in general for a \textit{holy} will: the “ought” is out of place here, because volition is of itself necessarily in accord with the law” (4:414). For Kant, the moral law is by itself sufficient to determine the power of choice. Therefore, absent the presence of a hindering impulse, the moral incentive would always be incorporated into the maxim over the incentives that are not moral. Insofar as a holy will is a will that admits no such hindering impulse, it is thereby insusceptible to the temptations these sensuous impulses bring, and so it is not constrained by the moral law. Hence, the holy will acts spontaneously in accordance with the moral law: “In other words, a propensity to good would consist in a kind of spontaneous preference for the impersonal requirements of morality over one’s own needs as a rational animal with a built-in desire for happiness. Since this preference must itself be based on a maxim and, therefore, consist in a settled policy, it might be inappropriate to characterize it as “spontaneous”. The point, however, is that for such an incentive,
the moral incentive would, as a matter of course, always outweigh the incentive of self-love” (Allison 1990, p. 155).

Allison then argues that it is impossible for finite beings like us to possess a holy will. The holy will always acts in accordance with the moral law, which is the reason why the moral law is not represented as an imperative to such a will. By contrast, a finite will is related to the moral law so:

However, if reason solely by itself does not adequately determine the will; if the will is exposed also to subjective conditions (certain incentives) that are not always in accord with the objective ones; in a word, if the will is not in itself completely in conformity with reason (as actually the case with human beings), then actions that are cognized as objectively necessary as subjectively contingent, and the determination of such a will in conformity with objective laws is necessitation (4:412-413).

The presence of these subjective conditions affect the finite will in their own way, independently of the strictures of the moral law. These incentives might determine the will in such a way that a will that acts on them would nonetheless be in complete conformity with the moral law. But it is also possible that these incentives would determine the will to act in a way completely contrary to the moral law. Therefore, while reason might say that to do or to omit something would be good,

---

78 We would do well to press Allison on the concept of a propensity to good. Now Allison claims that finite sensuous beings cannot possess a propensity to good. He reasons thus: “… Finite, sensuously affected rational beings such as ourselves are not only autonomous moral agents but also creatures of desire and inclination, which, as resting on natural causes, are neither completely in our control nor necessarily in agreement with the dictates of morality. Obviously, this does not mean that we are incapable of subordinating our sensuously based desire to moral considerations and, therefore, incapable of virtue. It does mean, however, that we are never beyond the possibility of temptation and the need for moral constraint. And this means we cannot be thought to have a propensity to good” (Allison 1990, p. 156). But an infinite being cannot possess a propensity to good either. For God does not desire. This is evident from the following passage: “All desires are either immanent or transient, i.e. either they relate to the very thing which has them and remain in this thing or else they relate to something which is external to the thing. But neither can be thought in a being of all beings. First, an all-sufficient being cannot have immanent desires, simply because it is all-sufficient. For every desire is directed only to something possible and future. But since God has all perfections actually, there is nothing left over for him to desire as a future possibility. But neither can God be represented as desiring something external to him; for then he would need the existence of other things in order to fulfill the consciousness of his own existence. But that is contrary to the concept of an ens realissimum” (28:1059). Yet a propensity is the subjective ground of the possibility of an inclination or habitual desire (6:29). Insofar as it is impossible for God to have desires, there can be no subjective ground in him for the possibility of a desire, and therefore no propensity (and no propensity to good). But then there is nothing that can possess the propensity to good. Allison’s propensity to good is an empty concept.
it says this to a will that does not always act on something because it is represented as good, and this is solely because it is a will that is susceptible to certain subjective conditions. As such, the moral law acts as a constraint on the finite will. And the necessitation of such a will by the moral law is called an imperative. For this reason, Allison thinks that a finite will, exposed in this way to subjective conditions, cannot be said to be holy, that is, without the possibility of any conflict between its inclinations and morality.\textsuperscript{79} In contrast to finite wills, a will that is holy does not contain the possibility of a conflict between its inclinations and morality because it is not sensuously affected: “… Finite, sensuously affected rational beings such as ourselves are not only autonomous moral agents but also creatures of desire and inclination, which, as resting on natural causes, are neither completely in our control nor necessarily in agreement with the dictates of morality” (Allison 1990, p. 156). Insofar as Allison considers having the propensity to good and the possession of a holy will to be one and the same thing, the fact that the will of a finite being cannot be holy must therefore entail that it does not have the propensity to good either.\textsuperscript{80}

\textsuperscript{79} In a footnote, Wood wonders indirectly whether this would mean that Jesus was afflicted with the human radical propensity to evil (Wood 2009, p. 158n). For Allison contends that the presence of the propensity to good in a human being entails having a holy will. And a holy will would have a spontaneous preference for the impersonal requirements of morality over the human being’s needs as a sensuous being: “It seems clear, however, that Kant thought human nature incapable of such a disposition, since it would mean that the will is beyond the need for rational constraint and thus beyond the twin thoughts of duty and respect for the law. This does not mean that Kant thought it impossible to subordinate self-love to duty, since this is precisely what the moral law requires. It is only that we cannot do so spontaneously, without, as it were, giving self-love a hearing. But if this is beyond the capacity of human nature and we must attribute to it a moral propensity of some sort, it follows on a rigorist view that we must attribute to it the contrary propensity to evil” (Allison 2002, p. 342). But the Synoptic Gospels clearly record Jesus as having experienced temptation, with two notable instances being the Temptation of Christ (Matthew 4:1-11 – see also Mark 1:12-13 and Luke 4:1-13) and at the Garden of Gethsemane (Matthew 26:36-45 – see also Mark 14:32-42). And if the experience of temptation means that the human being has not been able to spontaneously subordinate self-love to duty, then the portrayal of Jesus in the Synoptic Gospels show that He too has not been able to spontaneously subordinate self-love to duty. And if this is so, and we must attribute to Jesus a moral propensity of some sort, it follows on a rigorist view (if Allison is correct) that we must attribute to Him the contrary propensity to evil. Hence Wood observes: “So if we assume that these temptations arise from the radical propensity to evil, then we must conclude that this moral idea of humanity is also afflicted with that propensity. The holiness of will exemplified by the moral ideal consists not in immunity from the radical propensity to evil but rather only in not yielding to it” (Wood 2009, p. 158n).

\textsuperscript{80} Wood rejects the reasoning given here, arguing that Allison has reduced evil into “a trivial practical corollary of our finitude” (Wood 1999, p. 287). For the human being is denied the propensity to good because he is a finite sensuous being. And since the human being must have either a propensity to good or a propensity to evil, it follows that the human being is ascribed the propensity to evil because he is a finite sensuous being. Evil therefore becomes
But, Allison goes on to argue, Kant is an *ethical rigorist*, and this means that human nature must either be good or evil. There can be no middle position between the two extremes:

Now, 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) (6:24).

It is by appeal to this ethical rigorism that Allison intends to prove that the propensity to evil can be ascribed to the entire human race. Because if the will of a finite being is not that of a holy will, this has to be because the deviation from the moral law had been incorporated into his maxims, which prevents it from being so. And because no finite being has a will that is holy, this has to be because every finite being has incorporated the same deviation from the moral law into his maxims, which prevents their wills from being so. Thus, the propensity to evil is *universal*.

Allison’s proof does not end here, however. In what amounts to a separate sub-proof, we are given Allison’s explanation as to why his formal proof of the propensity to evil amounts to a synthetic proof. The reasoning he gives in defense of this is as follows: “And since, as we shall

---

a byproduct of our being finite creatures, thereby incurring the charge that it is a trivial practical corollary of our finitude. But this is obviously incorrect. There is no evidence at all that Kant thought that evil could be reduced to something so trivial. Indeed, Kant insists that evil is “an invisible enemy, one who hides behind reason and hence all the more dangerous”, which would hardly seem to describe human finitude. Furthermore, Wood points out that the mere fact that we are finite beings is not sufficient for the ascription of the propensity to evil to us: “… From the finitude of a will (or the presence to it of desires other than those prompted by pure reason) it simply does not follow that it is not holy (at most it follows that, for all we know, it *might* not be holy, because those desires *might* be contrary to reason and the will *might* prefer them as incentives over the incentives of reason). Even if we did know that the human will is not holy, this would tell us only that it does not *necessarily* follow the moral law, not that it displays a propensity not to follow the law. Still less could we infer from either the finitude or the nonholiness of the human will that there is in human nature a *universal, innate* propensity of this kind, which lies at the ground of all the evil deeds human beings do” (Wood 1999, p. 402). This brings to mind the possibility of *moral luck*. Allison says that the finite being is such that the moral incentive does not always outweigh the incentives of self-love, hence he would always be tempted to adopt maxims that run counter to the law and, therefore, regard the law as constraining (Allison 1990, p. 155). But is it not possible for a finite being to have the good fortune of living a life in which the moral incentive would, as a matter of course, always outweigh the incentive of self-love, and therefore never be tempted to adopt maxims that run counter to the law and, therefore, never think of the law as constraining?
see, the impossibility that is at issue is not logical (the notion of a propensity to good is not self-contradictory for Kant), the conclusion has a synthetic *a priori* status” (Allison 1990, p. 155). For Allison, what makes the conclusion that finite beings cannot have the propensity to good a synthetic proposition is the fact that the propensity to good is not self-contradictory. For suppose the concept of a propensity to good were to indeed be self-contradictory, then it would be true that finite beings like us cannot have such a propensity. But this would be an analytic truth having to do with the fact that anything that is self-contradictory would be logically impossible. Insofar as nothing exists that is logically impossible, it follows trivially that a property that is logically impossible cannot be instantiated in any existing being. Because he claims that the holy will that is necessary in order to have a propensity to good is unobtainable for finite beings like us, however, Allison thinks that this makes his conclusion synthetic: “As the text also makes clear, the problem with these positions is not that the concept of such a higher form of morality is self-contradictory (which would make the claim that finite rational agents cannot possess such a disposition analytic); it is rather that it reflects an ideal of moral perfection that is unobtainable by finite, sensuously affected agents such as ourselves” (Allison 1990, p. 156). And so Allison considers his formal proof of the universality of the propensity to evil to be synthetic solely because the propensity to good is possible for beings other than us, insofar as a propensity to good may be ascribed to God, even though he admits that it is impossible for finite beings like us to act without at least giving our desires and inclinations a hearing (Allison 1990, p. 156).

But is Allison’s formal proof for the propensity to evil the synthetic *a priori* proof he claims it is, or is it actually an analytic *a priori* proof in disguise? Now Kant himself denies the analytic *a priori* status of radical evil. He clarifies what he means when he says that the human being is by nature evil, explaining that: “… ‘He is evil by nature’ simply means that being evil applies to him
considered in his species; not that this quality may be inferred from the concept of his species (from the concept of a human being in general, for then the quality is necessary), but rather that, according to the cognition we have of the human being through experience, he cannot be judged otherwise, in other words, we may presuppose evil as subjectively necessary in every human being, even the best” (6:32). The formal proof that Allison has proposed seems to run afoul of this stricture. For if anything can be inferred from the concept of a human being in general, it seems that finitude would be something that belongs to this category of things, and the conclusion that the propensity to evil may be presupposed as necessary in every human being follows from this. This is because Kant considers the concept of finitude a limit on the infinite, as evidenced by the following passage:

… All three concepts, however – that of an *incentive*, of an *interest* and of a *maxim* – can be applied only to finite beings. For they all presuppose a limitation of the nature of a being, in that the subjective constitution of its choice does not of itself accord with the objective law of a practical reason; they presuppose a need to be impelled to activity by something because an internal obstacle is opposed to it. Thus they cannot be applied to the divine will (5:79).

This passage makes it clear that the hindering impulses that a finite being experiences are the result of the limitations of its nature. Insofar as the nature of an infinite being does not have any such limitations, such a being would not have needs that require satisfaction, since needs are deficiencies that would be inconsistent with the nature of an infinite being. This means that, if Allison is right that merely having hindering impulses prevents our will from being holy, then the opposition that prevents a finite being from having a holy will must involve a logical opposition and not a real opposition. For a finite being is, by definition, a being that has certain limitations on its nature that are not found in an infinite being. If, *per impossibile*, there is a finite being whose will is not subject to the hindering impulses that Allison describes, then this would mean that the limitations of its nature from which these internal obstacles derive are no longer present, and this
finite being would turn out to be an infinite being after all, which is absurd. And if, per impossibile, there is an infinite being whose will is subject to the hindering impulses that Allison describes, then this would mean that the nature of such a being is limited by internal obstacles, and this infinite being would turn out to be a finite being after all, and this too is absurd. Thus, it is a logical impossibility for a finite being to have a holy will under Allison’s account. As such, Allison’s conclusion that the propensity to evil is universal to finite beings like us must be regarded as an analytic proposition.

This result gives us grounds to be suspicious of Allison’s formal proof. The first suspicion is this: If Allison is right, and the propensity to evil can be ascribed to us on the basis of our finite nature, this has grave consequences for Kant’s doctrine of the immortality of the soul. Recall Kant’s argument for the immortality of the soul. The production of the highest good in the world is the necessary object of a will determinable by the moral law. Insofar as the complete conformity of the will with the moral law must be assumed for the production of the highest good in the world, the complete conformity of the will must be possible if the production of this highest good is possible. Because complete conformity of the will with the moral law is holiness, and this is a perfection that no rational being of the sensible world is capable at any moment of his existence,

---

81 If holiness is the goal for every human being, then there must be a way for a human being to achieve this goal. Now Kant describes how we are to achieve this holiness in the *Metaphysics of Morals*, in his discussion of our duty of self-perfection: “A human being has a duty to carry out the cultivation of his will up to the purest virtuous disposition, in which the law becomes also the incentive to his actions that conform with duty and he obeys the law from duty. This disposition is inner morally practical perfection” (6:387). The human being has a duty to cultivate his will, and the more he cultivates his will, the better he is able to withstand the impulses of nature that impede the moral disposition in him (6:380). The perfection of the human being’s will would ensure that he is always able to withstand the impulses of nature that impede the moral disposition in him. This represents a spontaneous preference for the impersonal requirements of morality over the human being’s needs as a rational animal with a built-in desire for happiness, which Allison calls holiness or the propensity to good (Allison 1990, p. 155). Holiness is therefore virtue (or human virtue in its most perfect form). This makes holiness an end that is also a duty (6:383). We ought to be holy, but if Allison is right, as sensuous finite beings, we cannot be holy. Therefore, if Allison is right, the principle that ought implies can must be given up in light of the reality of human finitude, because we can either judge that we can do what the law tells us unconditionally that we ought to do (6:380), or judge that the finitude of the human condition destroys any possibility of adopting the virtuous disposition that is our duty (6:383).
there is an infinite distance between the finite will and the holy will. Hence Kant posits the endless progress of the will towards complete conformity with the moral law as a substitute, and with this endless progress of the will there must be an endless endurance of the will to accompany it, which grounds our belief in the immortality of the soul. But this entire argument falls apart if holiness is something that is logically incompatible with the finite will. For an infinite distance between the finite will and the holy will presupposes a continuum that exists between the two points, even if it is assumed that traversing all the points in between is really impossible. This becomes untenable if the finite will is incompatible with holiness, and not merely incapable of holiness, because then the line between the finite will and the holy will would lose its end-point, and there would be nothing for the will to endlessly progress to. The endless progress of the finite will towards holiness is therefore impossible in precisely the same way the finite will is incapable of being holy, that is, logically impossible according to Allison. And while the primacy of practical reason gives us license to believe in things which practical reason finds necessary but speculative reason is silent on, it is not allowed to contradict speculative reason, which it would have to do in order to maintain its belief in the immortality of the soul (5:120).

The second suspicion that emerges from this line of reasoning is this: If Allison is right, and the propensity to evil can be ascribed to us on the basis of our finite nature, then Kant was deeply confused when he declared that human beings are by nature evil, or at least ignorant of the full implications of his own argument. For if Allison is right, every rational being other than God must be ascribed a propensity to evil on account of their finite natures. But would Kant really be willing to go so far as to make the bolder claim? It seems like he would hesitate to say that finite beings are by nature evil. For consider the case of rational beings other than humans, and whether

---

82 Indeed, it seems that Allison is simply wrong to say that human beings cannot possess a propensity to good. When he tries to determine what a propensity to good would look like in a human being, he writes: “… A propensity
or not such beings would be marked by a propensity to evil. Insofar as finitude is a characteristic that cuts across the phenomenal-noumenal divide, it seems reasonable to consider the possibility of finite beings that are not embodied. Human beings, as finite beings that are embodied, have to be impelled to an activity by the moral law because of an internal obstacle that is opposed to the moral incentive, an internal obstacle that only exists because of its sensible nature: “Impulses of nature, accordingly, involve obstacles within the human being’s mind to his fulfillment of duty and (sometimes powerful) forces opposing it, which he must judge that he is capable of resisting and conquering by reason not at some time in the future but at once (the moment he thinks of duty): he must judge that he can do what the law tells him unconditionally that he ought to do” (6:380).

Here we see Kant connecting the obstacle to practical reason to sensibility, the presence of which also entails finitude. But he also entertains the possibility of beings that are free from all sensibility, in whom sensibility cannot be an obstacle to practical reason. Insofar as these beings that are free from all sensibility are distinguished from the supreme being, they must have finite natures, yet it does not follow from this that beings that are free from all sensibility are marked by the same propensity to evil that we human beings are. For all we know, there might be angels, whose wills conform unhesitatingly with the divine law, despite the limitations of their natures. In conclusion,

to good would consist in a kind of spontaneous preference for the impersonal requirements of morality over one’s own needs as a rational animal with a built-in desire for happiness. Since this preference must itself be based on a maxim and, therefore, consist in a settled policy, it might seem inappropriate to characterize it as “spontaneous”. The point, however, is that for such an agent the moral incentive would, as a matter of course, always outweigh the incentive of self-love. Consequently, for an agent blessed with such a propensity, there would be no temptation to adopt maxims that run counter to the law and, therefore, no thought of the law as constraining” (Allison 1990, p. 155).

Now Allison’s argument is that human beings cannot have a propensity to good because they are finite sensuously affected beings (Allison 1990, p. 155). From this, it follows that finite sensuously affected beings cannot have a spontaneous preference for the impersonal requirements of morality over their own needs as rational animals with a built-in desire for happiness. Yet Kant explicitly considers finite beings that would never be tempted to adopt maxims that run counter to the moral law in the following passage: “For finite holy beings (who could never be tempted to violate duty) there would be no doctrine of virtue but only a doctrine of morals, since the latter is autonomy of practical reason whereas the former is also autocracy of practical reason, that is, it involves consciousness of the capacity to master one’s inclinations when they rebel against the law, a capacity which, though not directly perceived, is yet rightly inferred from the moral categorical imperative” (6:383).
when Allison says that finite, sensuously-affected beings cannot be assigned the propensity to good, and must therefore be assigned the propensity to evil, it is not altogether clear whether human beings cannot be assigned the propensity to good because they are finite or because they are sensuous, despite Allison’s assumption that it is due to our finitude when he extends his argument to include all finite rational beings (Allison 2002, p. 342).

Allen Wood

Lastly, we arrive at the interpretation of radical evil proposed by Wood. He draws upon the idea that the doctrine of radical evil in the Religion is based on Kant’s anthropology, following an earlier suggestion from Sharon Anderson-Gold. According to Wood, radical evil arises in the social context, and is to be identified with unsociable sociability, which is defined in the following passage from the Idea for a Universal History from a Cosmopolitan Point of View: “The means that nature employs in order to bring about the development of all the predispositions is their antagonism in society, insofar as this antagonism ultimately becomes the cause of a law-governed organization of society. Here I take antagonism to mean the unsociable sociability of human

---

83 For a different anthropological take on Kant’s doctrine of radical evil, see the development model that has been proposed by Mariña. According to the Lectures on the Philosophical Doctrine of Religion and the Conjectural Beginnings of Human History, the human being in his original state was determined by his animalistic impulses: “So long as human beings live only in the moment, guided by instinct alone, they are in paradise. Because at this point in their development reason does not yet exist in them and thereby makes no demands, they also incur no guilt” (Mariña 2017, p. 187). At this stage of human development, reason has not yet entered the picture. And when it finally does enter the picture, it is unable to contend with the fully-developed animal nature in the human being: “Reason, however, does not first come on the scene in its full maturity and strength, but only in its infancy, its powers weak. At first it is always the worsted party in its “scuffle with animality in its whole strength”; it is to be expected that its first exercise will be almost fully bungled, for along with reason come fresh desires due to the expanded powers of the imagination, and hence a new array of possible mistakes. There is just too much to manage” (Mariña 2017, p. 188). Yet there can be no moral evil before reason enters the picture. Prior to reason, the human being simply acted on his animal instincts, which were all he had. He was no more guilty of his savagery than a lion. But once reason enters the picture, weak as it is, the human being is confronted with a choice as to which he should follow. Even though the immaturity of his faculty of reason leads him to inevitably side with his animal instincts, the human being is morally responsible for this choice. Hence: “… Once reason begins to stir, the person no longer simply is what she finds herself to be. Reason demands that she must make herself into what she is to become. Since reason represents a moral demand and hence a choice, she becomes responsible for persisting in her condition of animality insofar as she does so, for she has chosen to identify with her animal needs. In this way has she made herself into an animal” (Mariña 2017, p. 187).
beings, that is, their tendency to enter into society, a tendency connected, however, with a constant resistance that continually threatens to break up this society” (8:20). Unsociable sociability is therefore Kant’s name for this pair of tendencies that work at cross purposes with each other. According to Kant, there is a sociable tendency that is inherent to human nature. As such, human beings have an innate inclination to associate with one another, in order to develop their natural predispositions, per Kant’s account of universal history (8:21). But Kant believes that there is also an unsociable tendency that is inherent in human nature. As such, human beings also have a tendency to separate themselves from society: “But they also have a strong tendency to isolate themselves, because they encounter in themselves the unsociable trait that predisposes them to want to direct everything only to their own ends and hence to expect to encounter resistance everywhere, just as they know they themselves tend to resist others” (8:21).

The universality of radical evil can therefore be explained by its origins in the predispositions of animality and humanity. Now the original predisposition to animality is described in the following passage: “The predisposition to animality in the human being may be brought under the general title of physical or merely mechanical self-love, i.e. a love for which reason is not required. It is threefold; first, for self-preservation; second, for the propagation of the species; third, for community with other beings, i.e. the social drive” (6:26). It is this social drive that we get from the predisposition to animality that constitutes the sociable aspect of unsociable sociability. The original predisposition to humanity, on the other hand, is described so: “The predispositions to humanity can be brought under the general title of a self-love which is physical and yet involves comparison (for which reason is required); that is, only in comparison with others does one judge oneself happy and unhappy” (6:27). Insofar as the human being is a rational animal, both original predispositions must be found in him as belonging with necessity to the human being,
because without the predisposition to animality, the human being would be rational without being an animal, and without the predisposition to humanity, the human being would be a mere animal incapable of rational thought. In neither case would the human being be a rational animal. But Kant also writes with respect to the predisposition to humanity: “... Out of this self-love originates the inclination to gain worth in the opinion of others, originally, of course, merely equal worth: not allowing anyone superiority over oneself; bound up with the constant anxiety that others might be striving for ascendancy; but from this arises gradually an unjust desire to acquire superiority for oneself over others” (6:27). And it is this constant anxiety that is responsible for the competitiveness of human nature, without which there would be no propensity to evil, but without which there also would be no progress (Wood 1999, p. 289).

How does competitiveness lead the human being to evil? The moment a human being finds himself in the company of others, his predispositions (or unsociable sociability) determine him not to allow anyone else gain superiority over himself, and this places him in a state of constant anxiety in which he suspects everyone around him of striving for ascendancy over him. This warps his perspective towards those around him. Instead of seeing them as potential allies, he sees them as competitors and potential threats. This competitive spirit is, according to Wood, in direct conflict with the basic requirements of the moral law (Wood 2009, p. 159). The moral law tells us “not to make an exception of ourselves to maxims we will hold as universal laws, to treat all rational beings as ends in themselves rather than subordinating them to our ends, to follow the laws of a realm of ends, in which human ends are in systematic harmony” (Wood 2009, p. 160). The competitive spirit tells us to do precisely the opposite. If we treat all rational beings as ends in themselves rather than subordinating them to our own ends, we make ourselves vulnerable to the machinations of others, thereby all but giving our competitors superiority over ourselves. This
gives rise to diametrically opposed standpoints. The moral standpoint tells us to follow the laws of a realm of ends, in which it is assumed that human ends can be put in systematic harmony. By contrast, the competitive standpoint considers human ends to be in contention with each other, and tells us to follow the counsels of prudence instead.\textsuperscript{84} Hence Wood writes: “Once we see that our natural inclinations, when shaped by our social conditions as rational beings, involve this competitive spirit, then we can see that the fundamental maxim of evil, which gives their satisfaction priority over obedience to the moral law, is really nothing except the desire for superiority over others and a policy of esteeming ourselves on the basis of our state or condition, which can be compared with that of others with the aim of validating that superiority” (Wood 2009, p. 160).

But are we morally responsible for this state of unsociable sociability that we find ourselves in? If the propensity to evil is identical with our unsociable sociability, then moral responsibility becomes problematic for Wood’s interpretation of the propensity to evil as unsociable sociability.\textsuperscript{85}

\textsuperscript{84} Wood addresses this objection in a way in the following passage: “… But it is one thing to say that the social condition provides the necessary context for developing our radical propensity to evil and quite another to say that society forces us to choose evil maxims, removing or diminishing our responsibility for these choices” (Wood 2009, pp. 168-169). For Wood, the bad social conditions a human being finds himself in play a necessary role in our choice of evil maxims, but does not absolve him of the moral responsibility he incurs for doing them, because he always has the freedom to refrain from making evil choices (Wood 2009, p. 169). The incorporation thesis is instructive here: the same way freedom of the power of choice cannot be determined to action through any incentive except so far as the human being has incorporated it into his maxim, so too are the social conditions in which the human being finds himself unable to determine the adoption of any incentive without the concurrence of the human being. But while the social conditions that play a necessary role in the human being’s choice of evil maxims do not exculpate the conduct that results from them, he cannot be blamed for the social conditions that are a necessary factor in his bad conduct, given that he is born into a situation in which human beings “mutually corrupt each other’s moral disposition and make one another evil” (6:94). And this is problematic because Kant considers the human being morally responsible for not just his bad conduct, but the evil propensity from which all this bad conduct issues (6:31). As Grenberg observes: “Wood’s assertion that radical evil, despite its social source, has a purely individual culpability is also compromised: why should I consider my tendency to place concerns of self above concerns of morality something for which I am ultimately responsible when, without the corrosive influence of other persons, my desires are “moderate”, “trquil”, and “undemanding”? ” (Grenberg 2005, p. 35). If society makes me into a villain, I am not responsible for my villainous nature, even if I am responsible for the villainies resulting from this nature.

\textsuperscript{85} Grenberg raises an objection on this point to Wood’s account of radical evil in \textit{Kant and the Ethics of Humility}, where she writes:
For Kant argues in the *Religion* that the propensity to evil cannot have an origin in time because it is a *moral* propensity, as opposed to a *physical* propensity (6:39). Moral propensities, Kant writes, pertain to a human being’s power of choice as a moral being, whereas physical propensities pertain to a human being’s power of choice as a natural being (6:31). The propensity to evil has to be a moral propensity because it has its origins in our freedom, which would not be possible if it is a physical propensity. Hence: “In [a physical propensity], there is no propensity to moral evil, for

The biggest problem for Wood is this: why is it that human beings, when in the presence of others, tend toward this fear of their own worthlessness? Why not just maintain one’s happy and contended state in a more communal setting? One could, perhaps, appeal to a quasi-Rousseauian account of scarcity of resources to explain the development of fear amongst otherwise content beings in a social situation. But this is not Wood’s, or Kant’s, story. It is, after all, not so much a fear for survival that is inspired on Kant’s account, but a fear of *worthlessness*. Appeal to the mere presence and recognition of other happy and content beings is similarly unsatisfying: happy and content beings encountering happy and content beings doesn’t seem to *demand* the development of fear and competition among these beings in a social situation. If some of these beings were to begin over-asserting themselves, Wood’s picture would follow. But why would even one happy and content being, with nothing more than the desire for equality, feel motivated to over-assert herself? Without a more extended reflection on the fact of human dependency, and the way it acts as a condition in this process, this question has no answer. All we can say is that beings who, on their own, would be happy and content, become unhappy, comparative, and competitive when put in a social situation; and that’s just the way humans are socially (Grenberg 2005, pp. 33-34).

In this passage, the issue with Wood’s account of radical evil as unsociable sociability is that there is a gap in the argument. For nowhere in the argument is it explained how the unjust desire to acquire superiority for oneself over others is able to gradually arise from the inclination to gain (equal) worth in the opinion of others: “… If everyone else has the desire for equal worth, and indeed that natural happiness and contentment of which Wood and Kant have spoken, the birth of this “constant anxiety” is unmotivated. It seems to come from nowhere. There is, then, a gap in this social story of the origin of radical evil: we cannot explain in social terms alone, why it is that fear and anxiety develop in social situations” (Grenberg 2005, p. 35). This presents a problem for Wood’s account of radical evil. In fact, it seems more pressing than Grenberg makes it out to be. If this fear and anxiety were absent, the human being never becomes *unsociable*, and there would be no unsociable sociability, and therefore no propensity to evil. Yet in searching for a *cause or reason* for this fear and anxiety, we are confronted with the troubling prospect that there is no suitable ground for this fear and anxiety but the *propensity to evil* itself, because they are vices that are grafted onto the inclination to gain equal worth in the opinion of others (6:27). Insofar as a propensity is defined as the subjective ground of the possibility of an inclination (6:29), and these *vices* of culture that are grafted onto competitiveness are *inclinations* (6:27), it follows that these vices (of fear and anxiety) are grounded on the propensity to evil. But then the propensity to evil cannot be explained by unsociable sociability, otherwise the argument becomes circular: the propensity to evil is the cause or reason of fear and anxiety, which is the cause or reason for the inclination to gain worth in the opinion of others becoming an unjust desire to acquire superiority for oneself over others, which is the cause or reason for the unsociable sociability in human beings that Wood identifies as the sole and sufficient explanation for the propensity to evil. Grenberg’s own solution gives us the same result. She writes that human finitude and dependence would fill the gap in Wood’s argument: “… Although inclinations themselves and the fact of human dependency are not evil, surely it would make sense to expect unsocial sociability, that fear of one’s own worthlessness in the sight of others, to find its footing in the propensities of an individual finite being. But Wood provides no such account and, as such, his story of the development of radical evil is lacking” (Grenberg 2005, p. 35). The point is that unsociable sociability presupposes something else, in this case finitude and human dependence, and therefore cannot be the ultimate cause or reason for the human propensity to evil.
the latter must originate from freedom; a physical propensity (one based on sensory inducements) to whatever use of freedom, be it for good or evil, is a contradiction. Hence a propensity to evil can only attach to the moral faculty of choice” (6:31). Later on, Kant goes over the same argument once again, but this time in terms of the distinction between an origin according to time and an origin according to reason (6:39). If the propensity to evil had an origin in time, its determining grounds lie in the state that precedes it in time. This would mean that the propensity to evil would have a natural cause in an event in the world, thereby making the propensity to evil a physical propensity (6:39-40). Hence: “To look for the temporal origins of free actions as free (as though they were natural effects) is therefore a contradiction; and hence also a contradiction to look for the temporal origin of the moral constitution of the human being, so far as this constitution is considered as contingent, for constitution here means the ground of the exercise of freedom which (just like the determining ground of the free power of choice in general) must be sought in the representations of reason alone” (6:40). The problem with the unsociable sociability that Wood wants to ascribe to human beings is that it gives the propensity to evil a temporal beginning. It is, after all, a misuse of the original predisposition to humanity, and therefore cannot precede the human being’s entrance into society, which has a beginning in time. 86 Therefore, Wood’s

86 Wood addresses a closely-related objection to the idea that the propensity to evil lies in the unsociable sociability of human beings. The objection points out that not every instance of moral evil can be traced to social competitiveness (Wood 2009, p. 166). The sort of moral evils that are connected to social competitiveness involve violations of our duties to others. This leaves violations of our duties to oneself unexplained by a propensity to evil that is rooted in the social condition. In response, Wood writes: “We misunderstand Kant’s solution if we think that it requires claiming that every individual instance of evil directly involves social competition. In particular violations of duties to oneself as an animal being – cases of suicide, gluttony, or drunkenness, for instance – may have a social aspect or they may not. (I may get drunk or kill myself because I have been humiliated by my social rivals, but I may also violate the same duties from motives having nothing to do with social competition). The point is rather that all such violations fundamentally exhibit the propensity to value one’s state or condition more than one’s person, and Kant’s solution to the propensity problem is that social competitiveness is the sole and sufficient for that propensity – whether or not social competitiveness is directly involved in its manifestation in a given case of evil choice” (Wood 2009, p. 167). Wood is certainly correct to insist that the propensity to evil is the ground for both the violations of duties to others and the violations of duties to oneself. But his critics are right to suspect that there is something amiss with the idea that social competitiveness is the sole and sufficient explanation for the propensity to evil. If social competitiveness is the sole and sufficient explanation for the propensity to evil, then if there is ever a human being
unsociable sociability can only be a physical propensity, not a moral propensity, if it can even be a propensity at all. And if Wood’s unsociable sociability can only be a physical propensity, and not a moral propensity, then we cannot be morally responsible for it.

The thrust of this objection, therefore, is that Wood’s interpretation of radical evil as unsociable sociability is inconsistent with Kant’s doctrine of freedom. Insofar as the propensity to evil has to have an origin in reason rather than an origin in time, it has to be an intelligible deed, and the choice of this propensity must be an exercise of the human being’s transcendental freedom. This is inconsistent with Wood’s interpretation of radical evil as unsociable sociability because the human being that is corrupted by his social conditions is a phenomenal being belonging to the world of sense, not a noumenal being belonging to the world of understanding. As such, transcendental freedom and unsociable sociability belong to two completely separate worlds, and never the twain shall meet. But Wood dismisses this objection as being based on a misinterpretation of Kant’s doctrine of transcendental idealism, as he says in the following passage: “This objection is based, in my view, on some very fundamental errors about Kant’s treatment of the problem of freedom and the role of transcendental idealism in resolving it. The function of Kant’s idea that we might be free as members of the intelligible world is only to show that there is no contradiction in regarding our actions both as free and as subject to the causal mechanism of nature in the sensible world” (Wood 2009, p. 165). For Wood, the concept of transcendental freedom and the metaphysical position that it entails is supposed to be treated as a postulate, a useful fiction for casting reasonable doubt on a transcendentally realist metaphysics in which everything is causally.

That is not part of society, the propensity to evil must either be absent or be unexplained. Now there might be solid anthropological reasons to think that human beings are always part of society, but it is not impossible to think of human beings apart of society, and we cannot rule out that there might actually be human beings leading such a solitary existence. And if this is the case, social competitiveness cannot be regarded as the sole and sufficient explanation for the universal propensity to evil.
determined in accordance with the laws of nature. Accordingly, there is no need to reify it into a dualist metaphysics in which the things in themselves are actually believed to exist. Hence Wood writes: “In assessing Kant’s compatibilism, it may help to remind ourselves that his theory of timeless agency is put forward only as a means of exploiting the burden of proof in the free will problem, which falls on those who would show that freedom is incompatible with determinism. Kant is not positively committed to his theory of the case as an account of the way our free agency actually works” (Wood 1984, p. 99).

Furthermore, Wood goes on to reject the metaphysical position of transcendental idealism, as a positive account of the way things are, as implausible: “If what bothers us about Kant’s theory is that it seems too farfetched and metaphysical, then it may help at least a little to realize that once the theory has served Kant as a device for showing that freedom and determinism cannot be proven incompatible, he is just as content to dissociate himself from it and adopt a largely agnostic position on how our freedom is possible” (Wood 1984, p. 99). But what are the aspects of transcendental idealism that Wood considers too farfetched and metaphysical to be believable? It seems like the aspect of transcendental idealism that Wood finds most problematic is the way it appeals to the doctrine of timeless agency in order to defend freedom from the threat of causal determinism.

According to the doctrine of timeless agency, it is possible for beings that do not exist in time to

---

87 Pereboom raises an interesting objection to Wood’s claim that Kant only means to exploit the burden of proof in the free will problem (Wood 1999, p. 99). He points out that the burden of proof shifts depending on the seriousness of the situation: “Kant may indeed have established that our being transcendentally free cannot be ruled out. However, in a large range of cases, such as that of the malicious liar, or the last murderer on death row in the island society, the guidance that belief in our transcendental freedom would provide is more aptly described as on the side of the prosecuting attorney. The epistemic standard that the prosecuting attorney must meet is not merely that his claims cannot conclusively be ruled out” (Pereboom 2006, p. 564). To continue the legal analogy, in a murder case in which the death penalty is on the table, the prosecuting attorney had better be able to prove beyond all reasonable doubt that the accused is guilty (or that transcendental freedom is real). In a matter of such seriousness, Kant does not get to play the defense attorney, as Wood argues here: “Kant’s role regarding freedom is somewhat like a defense attorney’s role regarding his client. Because practical freedom is presupposed by morality, we may assume that freedom is innocent until proven guilty, that the burden of proof lies on those who would undermine our moral consciousness by claiming that we are not free” (Wood 1999, p. 83). It would, after all, be perverse to demand that the accused prove himself innocent in a murder case in which the death penalty is on the table.
act in order to bring about temporal events. For Kant, the world of sense is determined by the laws of cause and effect, such that every event is preceded by an event from which it follows without exception, according to a rule. Insofar as a world represents the totality of appearances standing in thoroughgoing causal relations with one another, it follows that there must be as many possible worlds as there are possible configurations of appearances standing in thoroughgoing causal relations with one another. The way a timeless choice acts as a cause for temporal events is by selecting a certain subset of possible worlds from which the actual world is drawn, and this is what happens when the noumenal self makes a timeless choice to select its intelligible character, insofar as this is a choice that is compatible with only a subset of possible empirical selves (Wood 1984, p. 91). Therefore, even though our empirical character acts in a way that is determined by the laws of nature, Kant maintains that we have the ability to do otherwise, because we would have ended up with a different empirical character, had we made a different timeless choice: “Of every one of my misdeeds, it is true that I would have left it undone had I made a different timeless choice. Hence it is in my power to leave any misdeed undone, despite the fact that in the actual world it follows inescapably from what preceded it in time” (Wood 1984, p. 91). This is the aspect

---

88 Pereboom proposes an alternative way in which the timeless choice made by the noumenal self is able to determine the way the empirical world ends up being. Instead of a timeless choice restricting the number of possible worlds from which the actual world can be drawn, he proposes a Molinist reading of Kantian idealism: “According to the position of Luis de Molina on divine providence, God knows, eternally, what every possible libertarian free creature would choose in every possible circumstance, and with this knowledge, God is able to direct the course of history with precision, partly in virtue of creating just those free creatures whose choices fit a preconceived divine plan. On a version of this Molinist view adopted to Kant’s idealism, God would reconcile noumenal transcendental freedom with phenomenal determinism by creating just those transcendentally free beings the appearances of whose free choices conform to the deterministic laws that God intends for the phenomenal world” (Pereboom 2006, p. 557). Because my timeless choice does not play any part in the selection of the actual world from a (restricted) subset of possible worlds, Walkers worry might thereby be avoided, for it would no longer be true that “I can be blamed for the First World War, and for the Lisbon earthquake that so appalled Voltaire. Gandhi is no less guilty than Amin of the atrocities of the Ugandan dictator” (Walker 1979, p. 149). Pereboom admits that there are many who would find a Molinist reading of Kant less than credible for the theological and idealistic presuppositions it makes (Pereboom 2006, p. 558). But he writes: “… For the purposes of this discussion I don’t want to set the standard of credibility too high, and it suffices that idealism and Molinism are not uncommonly accepted by people who have considered these views ably and seriously” (Pereboom 2006, p. 558).
of transcendental idealism that Wood considers implausible, as an “absurd metaphysical fantasy that as free agents we are locked away in little monastic cells somewhere up there in the noumenal world” (Wood 2009, p. 166).

The problem Wood has with the doctrine of timeless agency is that it seems in no way incompatible with temporal striving and moral progress. Wood’s reasoning in *Kant’s Compatibilism* is as follows:

... There is no place in Kant’s theory for the idea that I may resist some passion or inclination of mine tomorrow by struggling with it today, and by striving throughout the day to purify my motives and fortify myself for the crucial hour of decision. The problem is not that I cannot imagine myself having all the thoughts and performing all the actions that I think of as part of this process, for I certainly can. The problem is that I cannot think of them as connected parts of an exercise of agency through time. I can only think of them as results or products of timeless agency, and not as the actual exercise of it. In time, there are only *facta*; yet trying or striving is not a *factum* but a *facere*. It is this exercise of agency Kant’s theory will not allow me to conceive as a temporal process (Wood 1984, p. 98).

The basic idea here is simple. Let us consider a straightforward example of moral progress from bad to good, where the agent is bad at $t_1$, neutral at $t_2$, and good at $t_3$. For the agent to undergo moral progress, we cannot simply conceive of the agent in terms of merely having the moral predicates “bad”, “neutral”, and “good”, because what matters is not simply that the agent has these predicates, but also the way these predicates are connected to each other in the agent. It matters here that the predicate “neutral” comes before the predicate “good” and after the predicate “bad”, because an ordering where the agent is good at $t_1$, neutral at $t_2$, and bad at $t_3$ is not an example of moral progress but of moral regress. Now Wood does not think that it is impossible for a noumenal agent to timelessly bring about these predicates “bad”, “neutral”, and “good” (or the noumenal analogs thereof). But because a temporal process where one stage is the *condition* for the next stage seems to be incompatible with the notion of a timeless agency, he thinks it is impossible is for a noumenal agent to timelessly bring about the predicate “bad” *followed by the*
predicate “neutral” followed by the predicate “good” (or the noumenal analogs thereof). And if Kant applies the doctrine of timeless agency to human agents, then this leaves them completely at odds with the way they represent themselves as striving and progressing: “The absence [of trying or striving] surely is a problem for dedicated Kantian moral agents, however, who must think of themselves as struggling constantly with their unruly inclinations and striving throughout their lives to make the idea of duty the sufficient motive of every action” (Wood 1984, p. 98).

But it is unclear whether, in explaining his objection to the doctrine of timeless agency, Wood is elucidating Kant’s position on the matter, follies and foibles all, or if he is reinterpreting Kant on the matter, and excising the parts he finds too farfetched and metaphysical. Admittedly, we find Kant in the First Critique using the theory of timeless agency to show that freedom and determinism cannot be proven to be incompatible, without committing himself to the theory as a positive account of how our free agency really is. Hence we find Kant framing the issue in merely hypothetical terms in the Dialectic of the First Critique, asking “Is it not rather possible that although for every effect in appearance there is required a connection with its cause in accordance

89 To be fair, Wood’s position on noumenal agency might be better described as agnostic. His interest in freedom and its relation to morality is focused on Kant’s argument that freedom must be presupposed as a property of the will of all rational beings (4:447-448). Hence he writes: “… The argument for freedom we have just seen is even more basic than Kant’s arguments for incompatibilism. For they say that whatever we may or may not hold about the compatibility of freedom and natural causality, we must presuppose our own freedom, as the capacity to act under norms of reason, in order even to represent ourselves as competent to decide on rational grounds whether fatalism or compatibilism is true. Our agreement or disagreement with Kant’s incompatibilism therefore should make no difference to our acceptance of his argument that [freedom] is a necessary presupposition of all rational judgment” (Wood 1999, p. 178). This is why Wood is able to write, after raising his objections to the idea of timeless agency, the following: “In the end, solving the free will problem may not be a matter of “saving common sense” (for that may be quite hopeless). Rather the solution may be a matter of saving as much of it as we can, and especially of saving those parts of it which matter most to us. I believe Kant saw the situation in this way, and I suggest we may assume that he decided the temporality of our agency is the necessary ransom that must be paid to the free will problem if our high vocation as moral agents is to be preserved” (Wood 1984, pp. 100-101). And yet still be able to dismiss the objections to his account of radical evil from the standpoint of transcendental freedom so: “Nothing Kant says could justify ascribing to him the absurd metaphysical fantasy that as free agents we are locked away in little monastic cells somewhere up there in the noumenal world” (Wood 2009, pp. 165-166). And the reason is because he considers the metaphysical position of transcendental idealism considerably less central to the practical side of Kant’s philosophy than the practical freedom of the *Groundwork* or the progressive development of the human race of Kant’s anthropology.
with laws of empirical causality, this empirical causality itself, without the least interruption of its connection with natural causes, could nevertheless be an effect of a causality that is not empirical, but rather intelligible…” (A544/B572). But this isn’t the entire story. For in the Second Critique, Kant takes himself to have proven that the transcendental freedom of this doctrine of timeless agency to be more than merely possible, but also actual. This is the point of the deduction of freedom in the Faktum Text, in which the Reciprocity Thesis, which shows that the concept of a will determinable by the moral law and the concept of a free will reciprocally entail each other, is used to prove that the will is transcendently free. To do this, Kant appeals to the fact of reason, which the moral law provides us, and this fact proves that our wills are wills for which the mere lawgiving form of a maxim can serve as a law, and in so doing proves to us that our wills are free wills.90 This allows Kant to conclude: “… And the moral law thus determines that which speculative philosophy had to leave undetermined, namely the law for a causality the concept of

---

90 In his analysis of the deduction of the categorical imperative in the *Groundwork*, Schönecker focuses on the following passage: “But because the world of the understanding contains the ground of the world of sense and so too of its laws, and is therefore immediately lawgiving with respect to my will (which belongs wholly to the world of understanding) and must accordingly also be thought as such, it follows that I shall cognize myself as intelligence, though on the other side as a being belonging to the world of sense, as nevertheless subject to the law of the world of understanding, that is, of reason, which contains in the idea of freedom the law of the world of understanding, and thus cognize myself as subject to the autonomy of the will; consequently the world of understanding must be regarded as imperatives for me, and actions in conformity with these as duties” (4:453-454). In this passage, the world of understanding is identified as the ground of the world of sense and also the ground of the laws of the world of sense. The human being regards himself as a member of both the world of understanding and the world of sense. But because the intelligible character of the human being is the ground of the empirical character of the human being, even though the human being regards himself as a member of both the world of sense and the world of understanding, he nevertheless takes himself to be subject to the laws of reason. This gives us what Schönecker calls the *ontoethical principle*: “Kant argues for the validity of the categorical imperative as a moral law for sensual-rational beings with the superiority of the ontic status of the world of understanding. The human being as a thing in itself and hence the ‘eigentliche Selbst’ and its law is of higher ontic value than the human being as an appearance and this is why the law of the world of understanding (the moral law) is binding upon the human being (as a categorical imperative) who is a member both of the world of understanding and the world of sense” (Schönecker 2006, pp. 316-317). But if the noumenal agent is just an “absurd metaphysical fantasy” (Wood 2009, pp. 165-166), then there is no ontoethical principle. For there would be no sense in saying that the “human being as a thing in itself and its law” is of higher ontic value than “the human being as an appearance and the laws of sense” when the former is a mere figment of the imagination and only the latter is in any way real.
which was only negative in the latter, and thus for the first time provides objective reality to this concept” (5:47 – emphasis mine).

Nor does Kant seem to find the incompatibility of the doctrine of timeless agency with trying and striving to be as problematic as Wood makes it out to be. Wood himself admits in Kant's *Compatibilism* that: “In some writings after 1793, Kant seems to recognize that his theory cannot accommodate moral striving or moral progress, literally speaking. And he seems to want to employ the notion of our noumenal “disposition” or “attitude” (*Gesinnung*) as a sort of timeless analogue or substitute both for moral striving and moral progress” (Wood 1984, p. 98). Indeed, a prime example of this can be found in the *Religion*, where Kant writes:

If by a single and unalterable decision a human being reverses the supreme ground of his maxims by which he was an evil human being (and thereby puts on a “new man”), he is to this extent, by principle and attitude of mind, a subject receptive to the good; but he is a good human being only in incessant laboring and becoming; i.e. he can hope – in view of the purity of the principle which he has adopted as the supreme maxim of his power of choice, and in view of the stability of this principle – to find himself upon the good (though narrow) path of constant progress from bad to better (6:48).

From this, we can see that Kant considers the doctrine of timeless agency to be compatible with trying and striving after all. According to Kant's account in the *Religion*, the moral progress of our earlier agent as he goes from bad at $t_1$ to neutral at $t_2$ to good at $t_3$ takes place in the mode of sense. Through it, a gradual reformation of our empirical character takes place: “Virtue, in this sense, is accordingly acquired little by little, and to some it means a long habituation (in the observance of the law), in virtue of which a human being, through gradual reformation of conduct and consolidation of his maxims, passes from a propensity to vice to its opposite” (6:47). But in the mode of thought, this moral progress is not represented as a process in which he goes from bad at $t_1$ to neutral at $t_2$ to good at $t_3$. That is, in the mode of thought, any change in the intelligible character is not represented as a gradual reformation, but is instead represented as a revolution in
disposition (6:47). For this reason, Kant describes the gradual reformation of character in the mode of sense in terms of a unity in the mode of thought: “For him who penetrates to the intelligible ground of the heart (the ground of all maxims of the power of choice), for him to whom this endless progress is a unity, i.e. for God, this is the same as actually being a good human being (pleasing to him); and to this extent the change can be considered a revolution” (6:48). The moral bent of the human being’s fundamental maxim acts, in Wood’s words, as a sort of timeless analogue or substitute for the gradual reformation of moral striving or moral progress that takes place in the world of sense. And it is the human being’s possession of a good fundamental maxim in the intelligible world that grounds his endless progress from bad to better in the world of sense. Thus, Kant does not give up the positive account of free agency we find in the doctrine of timeless agency, but instead he reinforces it by providing an analogue or substitute for trying and striving in the noumenal realm. Rather, it seems that, because Wood has chosen to interpret the propensity to evil as unsociable sociability, he is forced to deny the possibility of its being an intelligible deed through the rejection of Kant’s moral metaphysics. But in doing so, Wood has parted ways with Kant – he is no longer an interpreter of Kant but a revisionist bent on remaking Kant in his own image.

**Conclusion**

What is the missing formal proof of the radical evil in human nature in the *Religion*? Every attempt that has been surveyed in this chapter invariably has the same form: it begins with some characteristic \( C \) that is universally present in human beings, and argues that the propensity to evil is universally present in human beings because of \( C \), and moral evil is universally present in human beings because of this propensity to evil. But every attempt surveyed in this chapter fails. Either this characteristic \( C \) is universally present in human beings, and the propensity to evil is universally
present in human beings because of $C$, in which case the attempt fails because it makes the propensity to evil *analytically necessary*, and not something the human being can be held accountable for (6:32). Or this characteristic $C$ is universally present in human beings, but $C$ does not guarantee the presence of the propensity to evil in human beings, in which case the attempt fails because it would make the propensity to evil *merely contingent*, and why the propensity to evil is *innate* remains unexplained. These failed attempts have all had the form of a *progressive* argument – they attempt to provide a *sufficient condition* for the universal propensity to evil in the characteristic $C$ (Ameriks 2003, pp. 60-61). But perhaps a *regressive* argument, in which the universal propensity to evil is given as a condition for the possibility of moral evil, might fare better (Ameriks 2003, p. 60).\footnote{It might be helpful to introduce a distinction from Kant’s Pre-Critical philosophy here: “An *antecedently* determining ground is one, the concept of which precedes that which is determined. That is to say, an antecedently determining ground is one, in the absence of which that which is determined would not be intelligible. A *consequentially* determining ground is one which would not be posited unless the concept which is determined by it had not already been posited from some other source. You can also call the former the reason why, or the ground of being or becoming, while the latter can be called the ground *that*, or the ground of knowing” (1:392). If it is said that there is evil in the world, and the condition for its possibility is a universal propensity to evil, then the propensity is established as the antecedently determining ground for the evil in the world. It is the reason for the evil’s *being* in the world. Conversely, this evil in the world is a consequentially determining ground for the propensity. It is how we know that there is a universal propensity to evil. Therefore, a *regressive* argument for the universal propensity to evil uses the existence of evil in the world as a consequentially determining ground. By contrast, in an argument like Allison’s, the finitude of human beings is used as the antecedently determining ground to establish the universal propensity to evil. It is the reason why the propensity can exist. And while the propensity is a consequentially determining ground for the finitude of human beings, Allison does not appeal to it to establish the fact that human beings are finite creatures, but to entirely other grounds. This is why Allison’s argument, and other arguments like it, are *progressive* arguments for the universal propensity to evil.} This argument need not take the form of an empirical generalization, as Wood conjectured in *Kant’s Moral Religion* (Wood 1970, pp. 219-226). It is not necessary to demonstrate that *every* human being deviates from the law in order to establish a universal propensity to evil in the human species. For the deviation need only be *possible* for a human being in order to establish an *actual* propensity to evil in the individual human being. And since deviation from the moral law is possible for *every* human being, an *actual* propensity to evil is established to be in *every* human being, and the propensity to evil is thereby demonstrated to be
The missing formal proof of the radical evil in human nature in the *Religion* was exactly where it purported to be all along, but we could not see it, because we wrongly believed that the argument in question could only have a *progressive* form, and dismissed without due consideration the argument with a *regressive* form as an empirical generalization.
CONCLUSION

This is an essay about radical evil. And the issue that occupies every essay about radical evil is the problem of radical evil. It is the problem that arises when we follow Kant in saying that radical evil is universal and that radical evil is a product of our free choice. For, the standard objection goes, these two characteristics are at odds with each other. If radical evil is a characteristic that is present in the entire human species, or indeed ascribed to the species as part of its nature (6:32), then radical evil no longer seems to be something the individual human being can have any say in choosing. But Kant cannot allow such a result to stand, because a radical evil that is not a product of our free choice cannot be something for which we are morally responsible. If radical evil is the product of our free choice, then it would be up to every human being to adopt or refrain from adopting radical evil, and the odds that the same choice would be universally adopted by every individual human being is infinitesimal. The solution that the problem of radical evil demands from us is not a metaphysical one. The logical incompatibility of natural necessity and freedom already has a solution in the metaphysics of transcendental idealism – Kant’s solution to this is to consider the empirical objects and the laws that govern them as appearances, rather than as things in themselves, so that an effect “can therefore be regarded as free in regard to its intelligible cause, and yet simultaneously, in regard to appearances, as their result according to the necessity of nature” (A537/B565). But instead, the problem of radical evil is this: given that the intelligible world contains the ground of the sensible world and its laws, the universal propensity to evil in human nature has to have an intelligible cause, in the form of a reason why every human being chooses as he does. The solution to the problem of radical evil is to make sense of the noumenal choice of evil in the context of Kant’s moral-psychological framework.
Yet the fact that the problem of radical evil is the primary issue surrounding radical evil can only mean that the nature of radical evil is not an issue. Kant describes radical evil in two ways. First as a propensity: “By propensity (propensio) I understand the subjective ground of the possibility of an inclination (habitual desire, concupiscentia), insofar as this possibility is contingent for humanity in general” (6:29). Then as fundamental maxim: “Now, the term “deed” can in general apply just as well to the use of freedom through which the [fundamental] maxim (either in favor of, or against, the law) is adopted in the power of choice, as to the use of which the actions themselves (materially considered, i.e. as regards the objects of the power of choice) are performed in accordance with that maxim” (6:31). And these descriptions are unproblematically used to refer to one and the same thing, radical evil. Radical evil is the propensity to evil, and this means the same thing as being the evil fundamental maxim of a human being. Kant is responsible for this attitude with respect to radical evil. Throughout the Religion, we find passages such as these, where the two terms are used interchangeably: “… We are only talking of a propensity to genuine evil, i.e. moral evil, which, since it is only possible as the determination of a free power of choice and this power for its part can be judged good or evil only on the basis of its maxims, must reside in the subjective ground of the possibility of the deviation of the maxims from the moral law” (6:29). But a propensity is the subjective ground of an inclination, in contrast to a fundamental maxim, which is the subjective ground of a maxim, and a maxim and an inclination are by no means the same thing. Prima facie, a propensity to evil and a fundamental maxim that is evil, as the subjective grounds of two distinct things, must themselves be two distinct things. Thus, the fact that the nature of radical evil is not an issue, despite being identified as both a propensity and a maxim, is an issue.
The issues this essay deals with can therefore be broken down into three theses:

1. The propensity to evil has been misidentified with the evil fundamental maxim.
2. Kant’s argument for the universality of radical evil has been misunderstood.
3. The nature of radical evil in the noumenal realm has been taken for granted.

The propensity to evil has been misidentified with the evil fundamental maxim. For the most part, the literature has followed Kant in unproblematically treating the propensity to evil and the evil fundamental maxim as one and the same thing. This is evident in the case of Allison, who writes in his *a priori* proof of radical evil the following: “… Since evil has already been located in the subordination within a maxim of moral requirements to those of self-love, it follows that by a propensity to evil must be understood the meta-maxim to order the incentives in just this way in the adoption of an agent’s first-order maxims” (Allison 2002, p. 340). The analysis of the propensity to evil takes place in metaphysical terms as an intelligible act (either for or against the law) through which the fundamental maxim is chosen. At the end of this reductive analysis, the only thing that remains is the evil fundamental maxim, and all talk of a propensity to evil refers to this evil fundamental maxim. The opposite is true in the case of Wood. He writes in his anthropological proof of radical evil: “Nothing Kant says could justify ascribing to him the absurd metaphysical fantasy that as free agents we are locked away in little monastic cells somewhere up there in the noumenal world. Even in the first *Critique*, intelligible freedom is explicitly described as an intelligible *faculty* that belongs to the human being *as an appearance* – hence as a part of nature or the world of sense, not a faculty belonging to a separate, noumenal being” (Wood 2009, p. 166). Instead, intelligible freedom is reduced to a mere posit, something that Kant considers impossible to actually disprove, but whose existence he was largely agnostic towards (Wood 1984, p. 99). Absent this intelligible freedom, there is no intelligible act (either for or against the law),
through which the fundamental maxim is chosen. At the end of Wood’s reductive analysis, the only thing that remains is a propensity to evil, and all talk of an evil fundamental maxim is for the sake of the metaphysical thought-experiment that is transcendental idealism.

The issue here is fundamentally an issue about the separation of worlds. The propensity to evil, as an innate disposition, does not pre-exist the original predispositions to good, but is instead grafted onto them, like a shoot grafted onto a tree (6:28). These original predispositions belong to the possibility of human nature (6:28). Without the predisposition to animality, the human being would lack an animal nature, and would no longer be a rational animal. And without the predisposition to humanity, the human being would lack a rational nature, and would no longer be a rational animal. The human being is an appearance. Animals, rational or otherwise, are not things in themselves. The human being has to be assigned a place in the sensible world, and so too the propensity to evil that attaches itself to him, because to do otherwise would be to extend our cognition of the human being beyond the boundaries of sense. But the fundamental maxim belongs to the intelligible world. It is produced through “an intelligible deed, cognizable through reason alone apart from any temporal condition” (6:31). If the fundamental maxim is brought into the sensible world, then the use of freedom (either in favor of, or against, the law) that brings it about would have to be an event in time, which takes place in accordance with the law of cause and effect. This would be to give the origin of evil in human nature an origin according to time (6:34). And this Kant denies: “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 of reason; it cannot be derived from some preceding state or other, as must always occur, whenever the evil action is referred to its natural cause as event in the world” (6:39).
If we consider the propensity to evil and the evil fundamental maxim as one and the same thing, then we either bring the propensity into the intelligible world or bring the maxim into the sensible world, in which case either “the cognition we have of the human being through experience” of a propensity to evil is possible apart from the appearances (it isn’t) or an origin according time could be given for the origin of evil in human nature (it can’t).

Kant’s argument for the universality of radical evil has been misunderstood. Kant provides “a multitude of woeful examples that the experience of human deeds parades before us” as proof that the human being is by nature evil. The purpose of these examples has been misunderstood by the literature on radical evil. A standard response is Allison’s, who wrote: “… But clearly, even if for the sake of argument one accepts Kant’s appeal to some rather selective anthropological evidence, the most that this evidence can show is that evil is widespread, not that there is a universal propensity to it” (Allison 1990, p. 154). These interpreters dismiss the anthropological evidence that Kant provides as a bad attempt at an empirical generalization. And since the anthropological evidence is only enough to prove that the propensity to evil is widespread, and not universal, the empirical generalization fails, and they take it upon themselves to come up with a formal proof that Kant has neglected to provide. A more sympathetic response would be Wood’s in Kant’s Moral Religion, in which he wrote: “When Kant says that man is evil ‘by nature’, he does not mean to explain evil, but only to point out the universality of evil in man. Kant thus looks for evidence supporting the claim that all men, without exception, exhibit a propensity to evil; and he finds such evidence in “the multitude of crying examples which experience of the actions of men puts before our eyes” (Wood 1970, p. 225). The younger Wood also interprets Kant’s proof that the human being is by nature evil as an empirical generalization from the anthropological evidence he provides (Wood 1999, p. 287). But he thinks that the
anthropological evidence is enough to prove that the propensity to evil is universal. This did not last very long. By the time of Kant’s Ethical Thought (1999), Wood had backed away from the position of his younger self, and joined the ranks of the interpreters who consider the anthropological evidence to only be enough to prove that the propensity to evil is widespread, but not that it is universal. The two approaches both assume that Kant provides the anthropological evidence in support of an empirical generalization. The problem is that the proof that the human being is by nature evil is not an empirical generalization at all.

Kant actually provides the anthropological evidence to disprove rival theories of human nature. These theories claim that human nature is good, contra Kant’s claim that human nature is evil. The Romantics followed Rousseau in claiming that human beings are good in the state of nature. And it is they who Kant is referring to when he says, “… We wish to draw our examples from that state in which many a philosopher especially hoped to meet the natural godliness of human nature, namely from the so-called state of nature” (6:33). The anthropological evidence Kant provides of the “unprovoked cruelty in the ritual murders of Tofoa, New Zealand, and the Navigator Islands” serve as counter-examples that disprove the theory that human beings are good in the state of nature (6:33). The Enlightenment thinkers contemporary to Kant took the opposite stance. They claimed that human beings are good in the civilized state. And it is they who Kant is referring to when he says: “If we are however disposed to the opinion that we can have a better cognition of human nature in its civilized state (where its predispositions can be more fully developed), we must then hear out a long melancholy litany of charges against mankind” (6:33). The anthropological evidence Kant presents of what he calls the “vices of culture and civilization” serve as counter-examples that disprove the theory that human beings are good in the civilized state (6:33). Kant’s point in presenting this anthropological evidence is to show that the
competing theories that claim that human nature is good are wrong. And because Kant is also an ethical rigorist, it follows from his ethical rigorism that human nature can only ever be either good or evil. Therefore, by providing anthropological evidence falsifying the competing theories that claim that human nature is good, Kant indirectly provides support for his own theory that human nature is evil. This argument is not airtight, because Kant might not have falsified every competing theory that claims that human nature is good with his anthropological evidence, but neither is this an empirical generalization. It is a deductive argument, along the lines of Popper’s principle of falsification (Popper 1959, p. 19).

Furthermore, the “missing” formal proof for the universality of the evil in human nature is not actually missing, but right where we’d expect it to be in the *Religion*. It makes use of the same anthropological evidence that has been dismissed as insufficient for proving the universality of the evil in human nature. Recall the definition Kant gives of a propensity: “By propensity (propensio) I understand the subjective ground of the possibility of an inclination (habitual desire, concupiscentia), insofar as this possibility is contingent for humanity in general” (6:29). A propensity to evil is therefore the subjective ground for the possibility of an evil inclination, by which Kant means the inclination to subordinate the moral incentive to other incentives that are not moral (6:36). Kant is not making the claim here that the inclination to subordinate the moral incentive to other incentives that are not moral is a logical possibility for human beings. Per *The only possible argument in support of a demonstration of the existence of God*, this would be the formal element of possibility, “namely, agreement with the law of contradiction, [which] is cancelled by that which contradicts itself” (2:79). He is claiming that the inclination to subordinate the moral incentive to other incentives that are not moral is a real possibility for human beings. It is the material element of possibility that Kant is after here, where something is possible only
because “it presupposes something real, whether it be one thing or many” (2:79). The propensity to evil, as the subjective ground for the possibility of an evil inclination, is that reality in the human being which must be presupposed for evil deeds to be metaphysically or nomologically possible for him. The propensity to evil is therefore the condition for the possibility of an evil deed. And because the anthropological evidence proves that evil deeds are possible for every human being, this subjective ground for the possibility of an evil deed must be actually present in every human being, which is to say that there is a universal propensity to evil in human beings. The anthropological examples are not insufficient evidence for an unpromising empirical generalization. They are what allow Kant to echo the words of an English parliamentarian and declare that moral evil is a real possibility for every man (6:38). And if this is true, then the propensity to evil must be universal, or as Kant rhetorically puts it: “If this is true (and everyone can decide by himself), if nowhere is a virtue which no level of temptation can overthrow, if whether the good or evil spirit wins us over depends on which bids the most and affords the promptest pay-off, then, what the Apostle says might indeed hold true of human beings universally, “There is no distinction here, they are all under sin – there is none righteous (in the spirit of the law), no, not one” (6:38-39).

The nature of radical evil in the noumenal realm has been taken for granted. In the *Religion*, Kant explains the way in which an incentive can be made to coexist with the absolute spontaneity of the power of choice (6:24). We are told that the human being naturally incorporates both the moral and the nonmoral incentive into his maxim (6:36). The material of every maxim is therefore identical (6:36). The moral bent of a maxim therefore has to do with its form, as Kant explains in the following passage: “Hence the difference, whether the human being is good or evil, must not lie in the difference between the incentives that he incorporates into his maxim (not in
the material of the maxim) but in their *subordination* (in the form of the maxim): *which of the two he makes the condition of the other*" (6:36). If he subordinates the nonmoral incentive, then his maxim is a *good* maxim. But if it is the moral incentive that is subordinated, then his maxim is an *evil* one. The problem is that there is almost nothing in this account of maxims that is applicable to our choice of the fundamental maxim. The fundamental maxim is produced through an intelligible deed, which makes it “cognizable through reason alone apart from any temporal condition” (6:31). It is therefore the complete opposite of our regular maxims, which are “sensible, empirical, given in time” (6:31). But every incentive besides the incentive of reason (the moral incentive) is empirical, and belongs to the sensible world. They are therefore absent from the intelligible world, which is where the intelligible deed that produces the fundamental maxim takes place. This means that it is *not* the case that the moral incentive and the nonmoral incentive are both present in the fundamental maxim. But absent the nonmoral incentive, there would be nothing under which the moral incentive can be subordinated, and because there *must* be a subordination of the moral incentive in order for a maxim to be considered as evil, the choice of a fundamental maxim that is evil becomes impossible as a result. Yet Kant insists that we are morally responsible for the evil of our nature, and this is only possible if we actually made the choice that determined our fundamental maxim as evil.

If there is such a thing as an evil fundamental maxim, its formation cannot involve any reference to sensible incentives that simply do not exist in the intelligible realm where the deed takes place. The only material we have to work with is the moral incentive, which we are bound to incorporate into our maxim, in accordance with the laws governing the intelligible realm (4:446). It follows that the mere act of incorporating the moral incentive into our fundamental maxim does not make it good, because otherwise an evil fundamental maxim would have to be
dismissed as an impossibility. Instead, there must be space for freedom in the act of incorporation itself, insofar as it is up to each of us to decide how we want the moral incentive incorporated into our respective fundamental maxims. If the moral incentive is wholeheartedly incorporated into the fundamental maxim, then the human being's commitment to the moral law is absolute, and it would be impossible for him to be swayed by incentives that are not moral when he enters into the sensible world. But if the moral incentive is grudgingly incorporated into the fundamental maxim, then the human being is committed to the moral law only insofar as he is lacking in alternatives, and once he enters into the sensible world, he casts aside the moral law in favor of satisfying his newfound sensibilities. The difference between these two cases lies in their moral strength. This moral strength is necessary to maintain an absolute commitment to the moral law. But even those who are unable to maintain their commitment to the moral law due to their weakness have moral strength to some degree or the other. As Kant asserts in the Second Critique: “... But ask him whether, if his prince demanded, on the pain of the same immediate execution, that he give false testimony against an honorable man whom the prince would like to destroy under a plausible pretext, he would consider it possible to overcome his love of life, however great it may be. He would perhaps not venture to assert whether he would do it or not, but he must admit without hesitation that it would be possible for him” (5:30). Kant intends for this passage to demonstrate an individual’s awareness of his own freedom, because he is aware of what he ought to do, and through it becomes aware of the power he has to do the right thing, even though his every inclination is to do the opposite. This is what virtue is: the capacity and considered resolve to withstand what opposes the moral disposition in us (6:380). And even the most reprobate human being has a little bit of virtue in him, as the above passage demonstrates.
Because the intelligible world contains the ground of the sensible world and its laws, the selection of an evil fundamental maxim in the intelligible world produces the propensity to evil in the sensible world. It does so by grafting itself to the phenomenal structures that belong with necessity to the human being. In his discussion of the predisposition to animality, which Kant describes as a mechanical self-love, we are told that this predisposition is responsible for our drive for self-preservation, for the propagation of the species through the sexual drive, and the instinct for the preservation of the offspring thereby begotten (6:26). These three can be corrupted to produce the vices of the savagery of nature, which “are called the *bestial vices of gluttony, lust and wild lawlessness*” (in relation to other human beings)” (6:27). These are vices of excess, in which the natural drives of the predisposition of animality are not properly reined in, but allowed to run wild. The choice of an evil fundamental maxim results in a lack of moral strength, without which the natural drives of the predisposition to animality cannot be controlled. This moral weakness leads to frailty, or “the general weakness of the human heart in complying with the adopted maxims” (6:29). And in his discussion of the predisposition to humanity, which Kant describes as a comparative self-love, we are told that “out of this self-love originates the inclination to gain worth in the opinion of others, originally, of course, merely equal worth: not allowing anyone superiority over oneself; bound up with the constant anxiety that others might be striving for ascendancy; but from this arises gradually an unjust desire to acquire superiority for oneself over others” (6:27). This represents the slide from self-love to self-conceit, which is what would happen naturally if morality is not in control. The choice of an evil fundamental maxim results in a lack of moral strength, without which the gradually arising of an unjust desire to acquire superiority for oneself would arise out of our comparative self-love. And this moral weakness with respect to the predisposition to humanity results in depravity, which “can also be called the perversity of the
Thus, the problem of radical evil is a problem because of the way the propensity to evil has been identified with the evil fundamental maxim. Because of it, we have difficulty even formulating our problem with radical evil. If what we mean by the problem of radical evil is “How do we know the propensity to evil is universal?”, then we must search for our answer in the sensible world and its laws. The answer to this problem of radical evil has to do with the moral evil that is a real possibility for every man. This possibility for moral evil has to be grounded in the appearances as a propensity to evil, and because every man can possibly commit moral evil, it follows that every man must actually possess this propensity to evil. The propensity to evil is therefore a synthetic a priori fact about our phenomenal natures, and our cognition of it stems from inquiring into the conditions of the possibility of the anthropological evidence that we have. But we can also mean by the problem of radical evil “Why do human beings choose the propensity to evil?”, and here we are asking about the reasons for human actions, and not the phenomenal conditions of human action. This takes us to the intelligible world that is the ground of the sensible world and its laws. In making the noumenal choice for his fundamental maxim, the human being has only the moral incentive for his material, and any free exercise of his power of choice has to take place within the bounds of this restriction. There is no way to avoid incorporating the moral incentive into his fundamental maxim, but the way we can choose to incorporate it is another matter entirely. Every human being incorporates the moral incentive, but it “takes” to a greater or lesser degree. And this determines the moral strength he has to resist tests to his moral character. The problem of evil is obscured by the fact that it is operating at both the phenomenal level and the noumenal level, and giving the moral-psychological reasons for the propensity to evil do
nothing to address the demand for the phenomenal conditions *necessitating* the propensity to evil that motivate the questions about its universality, and *vice versa.*
BIBLIOGRAPHY


Quinn, Philip L. "In Adam's Fall, We Sinned All". *Philosophical Topics* 16, no. 2 (Fall 1988): 89-118.


