Cracks in the Keystone:
Kant on the Incomprehensibility of Freedom

By

Tim Aylsworth

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy
(Philosophy)

at the

UNIVERSITY OF WISCONSIN-MADISON

2018

Date of final oral examination: 4/26/2018

The dissertation is approved by the following members of the Final Oral Committee:

James Messina, Assistant Professor, Philosophy
Steven Nadler, Professor, Philosophy
Alan Sidelle, Professor, Philosophy
Anat Schechtman, Assistant Professor, Philosophy
Hannah Eldridge, Assistant Professor, German

Table of Contents

Acknowledgements

I owe a debt of gratitude to many people for helping me with this project. First, I would like to thank my committee for everything they have done for me: Steve Nadler, Alan Sidelle, Anat Schechtman, and Hannah Eldridge. I am especially grateful to my advisor, James Messina, who read many early drafts of this dissertation and whose feedback enormously improved the final version. I would also like to thank Karl Schafer, Allen Wood, and Andrew Chignell for reading the fourth chapter and giving me helpful comments. I greatly appreciate everyone who was willing to discuss these ideas with me as I was developing them, most notably Hadley Cooney, Hayley Clatterbuck, and Jacob Krch. I want to thank my parents whose endless, loving support helped make this possible. Thanks to Carol Rainwater for coming all the way to Thailand to help out at a crucial time. I am thankful for all my friends who encouraged me along the way. And most of all, thanks to Katie and Iris Rainwater for giving me the motivation I needed to reach the finish line. This work is dedicated to them.

Preface

"Now, the concept of freedom, insofar as its reality is proved by an apodictic law of practical reason constitutes the *keystone* of the whole structure of a system of pure reason, even of speculative reason . . ." (Kant, *Critique of Practical Reason*, 5:3–4).

"[F]reedom insofar as it is the basis of morality *cannot be comprehended*, and insofar as it can be comprehended it cannot serve as the foundation of morality" (*Kraus's Review of Ulrich's* Eleutheriology, 8:458).[1]

Kant's concept of freedom has puzzled generations of commentators and given rise to some intractable controversies. Indeed, the exegetical and philosophical difficulties are so severe that his view has been described as "incoherent,"[2] "metaphysically preposterous,"[3] and "worthless"[4] — perhaps a valiant effort but ultimately "a hopeless failure."[5] To make matters worse, there is no consensus on what the view even is. Some readers, lending Kant a charitable hand, have jettisoned much of the theory's apparent baggage and suggested that it might be nothing more than the ontologically unambitious claim that when we engage in moral deliberation we are justified in adopting a standpoint from which we think of ourselves as free.[6] Others have defended a metaphysical interpretation, claiming that timeless noumenal choices somehow determine states of affairs in the phenomenal world. Proponents of this view maintain that every rational being has an intelligible character, ontologically distinct from its empirical counterpart, and they argue that

---

[1] See the Cambridge collection *Practical Philosophy*, 129. Cf. Mary J. Gregor's editorial note about the inclusion of Kraus's review in the Academy edition of Kant's work, *Practical Philosophy*, 121.

[2] Henry Allison, "Empirical and Intelligible Character in the *Critique of Pure Reason*," 1. Of course, Allison is an ardent defender of Kant's view of freedom. The context here is that he is responding to critics who think the view is incoherent. Similarly, Graham Bird says that Kant's account is "widely regarded as unacceptable or incoherent" (Bird, *The Revolutionary Kant*, 696).

[3] Onora O'Neill, *Constructions of Reason*, ix. The same is true of O'Neill. She also defends Kant's view and argues for a two-standpoint interpretation. She is noting the scholarly ambivalence concerning Kant's view: "Kant is revered for his unswerving defense of human freedom and respect for persons, and for his insistence that reason can guide action. He is also reviled for giving a metaphysically preposterous account of the basis of freedom" (ix).

[4] Jonathan Bennett, "Kant's Theory of Freedom," 102–12.

[5] Ralph Walker, *Kant*, *the Arguments of the Philosophers*, 148. Cf. Jonathan Bennett, *Kant's Dialectic*, 210–14.

[6] See, for example, Christine Korsgaard, *Creating the Kingdom of Ends*, ch. 6. Cf. Hilary Bok, *Freedom and Responsibility*. Some, like Bok, defend this more as a "Kantian" view and less as an interpretation of Kant.

this explains how an action in the empirical world could be a consequence of both an uncaused, spontaneous choice and deterministic laws of nature.[7]

In spite of the many disputes, the one thing that everyone seems to agree on is that whatever Kant is trying to say about freedom, it is fraught with problems. Many objections have been raised and many questions left unanswered. This should be cause for concern given the singular importance that Kant attaches to free will. For better or worse, freedom has the privilege of being the "keystone" in the arch of his critical philosophy. It occupies a unique conceptual space at the intersection of practical and theoretical reason,[8] somehow holding the entire edifice together. Morality, as Kant understands it, requires freedom as a necessary condition: it is the "*ratio essendi* of the moral law" (*KpV* 5:5) Freedom is the source of "inner value of the world."[9] The special status assigned to autonomy is what makes the good will "shine like a jewel" — a sapphire in the mud of heteronomy (*G* 4:394).

In his political philosophy, Kant argues that only freedom can legitimate or justify the existence of the state. It is "the only original right belonging to every man by virtue of his humanity." (*MS* 6:237). Anticipating Hegel, Kant suggests that freedom is the ultimate end of all human history and culture.[10] He even thinks that we are compelled to entertain the idea that all of

---

[7] See, for example, Eric Watkins, *Kant and the Metaphysics of Causality*, 343–61.

[8] There are other ideas (like our idea of God), which Kant considers to be within the domain of both theoretical and practical reason, but freedom is given a special status. He says, "But among all the ideas of speculative reason freedom is also the only one the possibility of which we *know* a priori, though without having insight into it, because it is the condition of the moral law, which we do know. The ideas of *God* and *immortality*, however are not conditions of the moral law but only conditions of the necessary object of a will determined by this law" (*KpV* 5:4). Kant thinks that a moral person must set for herself an ultimate end (the highest good), which gives her practical reasons for believing that there is an afterlife in which she can continue pursuing moral perfection and where God will reward virtue with happiness.

[9] This is from an ethics lecture from the 1780s known as the *Moral Mrongovius* lecture: "If all creatures had a faculty of choice bound by sensuous impulses, the world would have no value; the inner value of the world, the *summum bonum*, is the freedom to act in accordance with a faculty of choice that is not necessitated. Freedom is therefore the inner value of the world" (27:1482). Cf. Paul Guyer, *Kant on Freedom, Law, and Happiness*, 96.

[10] Kant says that history is "nothing other than the transition from the crudity of a merely animal creature into humanity, from the go-cart of instinct to the guidance of reason — in a word, from the guardianship of nature into the

nature is teleologically directed toward the human being — the only creature on the planet in possession of this supersensible faculty — as nature's "final end."[11] Henry Allison may be right when he says, "It is no exaggeration to claim that, at bottom, Kant's critical philosophy is a philosophy of freedom."[12]

Thus, when we consider the array of problems surrounding Kant's theory of freedom and the many philosophical puzzles that beset scholars who try to make sense of the view, it should come as a surprise that relatively little attention has been paid to his many warnings that freedom is, in some sense, incomprehensible. He emphasizes this point in several places, but it is stated clearly and definitively in the *Groundwork*:

> Philosophy must therefore assume that no true contradiction will be found between freedom and natural necessity in the very same human actions, for it cannot give up the concept of nature any more than that of freedom. Nevertheless, this seeming contradiction must be removed in a convincing way, *even though we shall never be able to comprehend how freedom is possible*" (*G* 4:456; emphasis added).

What are we to make of this claim? Kant spills a great deal of ink explaining the details of transcendental freedom, arguing for its compatibility with determinism, and highlighting its crucial significance for his practical philosophy. Why then would he confess that we will never be able to comprehend how freedom is possible?

The aim of my dissertation is to answer that question by offering a novel explanation and defense of Kant's take on the incomprehensibility of freedom, thereby filling a significant lacuna in the literature. I conclude that when it comes to theoretical speculation about freedom, the

---

condition of freedom" (*Conjectural Beginning of Human History*, 8:185). For an illuminating and original defense of the idea that the realization of freedom is (and ought to be) the final end of human history and culture, see Kristi Sweet, *Kant on Practical Life: From Duty to History*.

[11] This idea, which I discuss in chapter 5, features prominently in the *Critique of the Power of Judgment*, especially in the second half concerning teleology (5:434–47). Cf. *Teleological Principles*, 8:182. See also Guyer's editorial note 14 to the Cambridge Edition to the third *Critique* (390). See also Guyer, *Kant's System of Nature and Freedom*, ch. 11.

[12] Allison, *Kant's Theory of Freedom*, 1.

following things are, in principle, unknowable: (1) the exact nature of the relation between intelligible and empirical character and (2) determinate, substantive facts about the noumenal self and its free agency as a spontaneous first cause. Although the fact of reason "furnishes reality to a supersensible object of the category of causality," and thus establishes the reality (and *a fortiori* the real possibility) of freedom, Kant is adamant that this is "only for practical use" (*KpV* 5:6). Practical knowledge about freedom, which springs from our awareness that we are bound by the moral law, tells us *that* we are free, but it does not justify any speculation about *how* we are free.

It will become clear in what follows that my conclusion is rather contentious and that it has some wide-ranging implications for those interested in Kant's work (and in Kantian ideas more generally). If Kant believes that the details of transcendental freedom are beyond our comprehension, then there are some scholarly debates that he might regard as misguided — precisely the kind of dialectical transgressions that he warns us to avoid. Perhaps the most notable case is the question of moral motivation and the role of moral feeling. If I am right about his pessimism regarding our understanding of freedom, then it is Kant's position that we can never know the exact role that moral feeling plays when an agent acts out of respect for the moral law. Furthermore, freedom is occasionally invoked as a bargaining chip in the interpretive debate over transcendental idealism, but I believe the incomprehensibility of freedom undermines the legitimacy of some of these arguments.

The implications of my conclusion, however, are not entirely negative. On the positive side, if Kant is right that there are principled reasons why freedom is beyond our comprehension, then this would lift an undesirable explanatory burden off the shoulders of Kant's critical philosophy. It would be a boon for Kantians if they could maintain their lofty claims about the unique, elevated status of rational agency without being committed to a detailed (and implausible)

explanation of transcendental freedom. It would be especially nice if the reasons for the incomprehensibility thesis are indeed principled, as I claim they are, and it is not just special pleading: a case of Kant's practical interests compelling us to ask theoretical reason to make an ad hoc exception for the crown jewel of his moral philosophy.

In order to take on this task of explaining Kant's position on the incomprehensibility of freedom, several issues must be addressed first. I begin chapter 1 by delineating the different concepts of freedom at work in Kant's thought. Freedom plays different roles in his practical and theoretical philosophy, and these disparate interests of reason assign distinct epistemic statuses to the various concepts of freedom. The interpretation I offer in the first chapter will serve as the groundwork for the later chapters which will rely on the taxonomy and definitions I put forward. I also explore the historical background and development of Kant's position on freedom, as his view changes over time. In his earlier works, he was satisfied with a compatibilist concept of freedom and, like Wolff and Baumgarten, he did not think transcendental (i.e., libertarian) freedom was necessary for morality. But he later rejected that view. I explain the considerations that led him to that shift, as this will help inform some of the discussions that follow.

In chapter 2, I discuss the role that freedom plays in Kant's theoretical philosophy. In the Transcendental Dialectic of the *Critique of Pure Reason*, Kant argues that freedom is an indispensable idea of speculative reason, but he thinks it inevitably seems to run afoul of the causal determinism argued for in the Transcendental Analytic. I explain these arguments and I show how he thinks that transcendental idealism resolves the conflict between freedom and determinism while nevertheless leaving us in an agnostic position. Theoretical reason cannot secure the reality or even the metaphysical possibility of freedom. The resolution of the antinomy shows only that freedom does not conflict with the deterministic laws of nature and that it is not logically

impossible. By underscoring the limited nature of theoretical reason's commitment to transcendental freedom, the second chapter anticipates my central thesis about the incomprehensibility of freedom.

After Kant demonstrates that freedom cannot be ruled out by theoretical reason, he argues that practical reason is what ultimately compels us to believe that rational beings really do possess transcendental freedom. I offer a new interpretation of this argument in chapter 3. I think Kant's view has been largely misunderstood by commentators, and I highlight the importance of moral feeling as the missing link between our consciousness of the moral law and the claim that we have free will. For Kant, the fact that the pure faculty of practical reason is capable of giving rise to a feeling of respect is proof that the pure will has palpable effects in the empirical world.

In chapter 4, I present Kant's arguments for the claim that we will never be able to comprehend how freedom is possible. I believe that the incomprehensibility of freedom has several important, underappreciated implications for Kant's theoretical philosophy and for Kantian ethics and metaethics more broadly. The secondary literature on Kant's concept of freedom contains scandalously few discussions of the incomprehensibility thesis; its implications and the arguments behind it have been unjustly neglected. I provide a close textual analysis in order to discern the content of the incomprehensibility thesis, and I formulate an argument to show how Kant's conclusion follows from his epistemic commitments. I conclude the chapter by outlining some of the most notable consequences of my view.

I make some concluding remarks in chapter 5 as I briefly discuss the role of freedom (and its incomprehensibility) in Kant's final critical work — the *Critique of the Power of Judgment*. He says that one of his aims there is to bridge the "incalculable gulf" between nature and freedom. I claim that this framing of the project of the third *Critique* can be seen as confirmation of the

conclusions I defend in my dissertation. The incomprehensibility of freedom is an essential background condition of the problem that Kant is trying to solve. I explain what Kant takes the problem to be and how he hopes to address it by means of the faculty of reflective judgment. I focus in particular on the problem posed by the highest good. The incomprehensibility thesis shows us that we cannot comprehend the connection between pure practical reason and sensibility, but Kant's commitment to the highest good requires us to believe in the possibility of virtue being rewarded with happiness — a belief that entails such a connection. Finally, I conclude the project by summarizing the central points of previous chapters and highlighting the connections between them in an effort to demonstrate the continuity and consistency of Kant's thoughts on freedom and its incomprehensibility.

Chapter 1
The Many Faces of Freedom

1.1 Introduction

In the first *Critique*, freedom presents a problem: reason, Kant argues, is ineluctably drawn to seemingly contradictory conclusions. On the one hand, the concept of causality is a condition of the possibility of experience, and we know a priori that every event is caused in accordance with the laws of nature.[13] On the other hand, he believes that rational beings like us have free will in the incompatibilist, libertarian sense. This leads to an apparent contradiction — an antinomy of reason — as we are drawn to ostensibly incompatible beliefs about free will. Furthermore, when we engage in speculation about freedom we are tempted to make substantive claims about things in themselves in spite of the fact that we have no theoretical cognition (*Erkenntnis*) of the noumenal realm.[14] Without cognition, our idea of the intelligible self lacks determinant content, yet we speculate about the possibility of its freedom anyway.

Perhaps these difficulties are to be expected considering Kant's warning at the outset of the first *Critique* regarding the "peculiar fate" of reason which burdens it "with questions which it cannot dismiss, since they are given to it as problems by the nature of reason itself, but which it also cannot answer, since they transcend every capacity of human reason" (A vii). Freedom is a paradigmatic instance of reason's peculiar fate.[15] Kant thinks that we are compelled to ask

---

[13] As Kant puts it in the second edition, "All alterations occur in accordance with the law of the connection of cause and effect" (B232).

[14] It is important to distinguish theoretical cognition from practical cognition. As I explain in chapter 3, Kant argues in the second *Critique* that we *do* in fact have practical cognition of freedom even though one of the main themes of the first *Critique* is that we cannot cognize things in themselves theoretically.

[15] The other notable cases of reason's desire to overstep its bounds are God and the soul. One of Kant's principal aims in the first *Critique* was to chastise what he saw as "dogmatic metaphysics," especially with respect to its claims that we can have a priori knowledge of God, freedom, and the soul. Transcendent metaphysical speculation is not limited to these three things, however. There are other domains where reason speculates beyond the boundaries of possible knowledge. For instance, the first Antinomy deals with cosmological speculation about whether or not the world has a beginning in time or a limit in space.

unanswerable questions about freedom. And even after transcendental idealism rescues reason from the apparent incompatibility of freedom and determinism, we nevertheless end up in a state of uncertainty in keeping with Kant's epistemic humility. In the end, theoretical reason cannot demonstrate the real possibility of transcendental freedom; it can show only that it is not logically impossible and that it is not necessarily inconsistent with determinism.

Near the end of the first *Critique*, it becomes clear that Kant's metaphysical and epistemological views are an essential propaedeutic for his practical philosophy — where the centrality of freedom becomes even more salient.[16] For Kant, freedom and morality are two sides of the same coin. He says in the *Groundwork* that "a free will and a will under moral laws are one and the same" (*G* 4:447). From his earliest pre-critical works until his death, Kant remained unequivocal about the importance of freedom for his practical philosophy. Morality is paramount to Kant,[17] and he believes that, without freedom, there can be no moral law. It should come as no surprise that there is little scholarly dispute about the importance of the concept of freedom in Kant's work. There is general agreement about the broad outline of the view and its importance. There is widespread disagreement, however, when it comes to the details.

The first problem Kant scholars face is that freedom plays not one role but several, and Kant makes use of at least four concepts of freedom.[18] It is further complicated by the fact that

---

[16] In the *Groundwork* Kant is explicit about the role that theoretical philosophy plays in paving the way for the practical philosophy: "Hence, it is an indispensable task of speculative philosophy at least to show that its illusion about the contradiction rests on our thinking of the human being in a different sense and relation when we call him free and when we hold him, as a part of nature, to be subject to its laws, and to show that both not only *can* very well coexist but also must be thought as *necessarily united* in the same subject . . . This duty, however, is incumbent upon speculative philosophy **only so that it may clear the way for practical philosophy**" (4:456; boldface added). Kant's practical writings contain many similar declarations. Cf. *KpV*; 5:6; *G*; 4:451

[17] Moral and political concerns about freedom become Kant's principal focus that after the first *Critique* humbles reason's pretensions about coming to have definitive knowledge about things in themselves, God, freedom, and the immortality of the soul. Some, like Richard Velkely, take the primacy of the practical much further, arguing that the even the project of the first *Critique* is driven by moral ends. See Richard Velkley, *Freedom and the End of Reason*.

[18] The final number ultimately depends on how the concepts are counted. Lewis White Beck finds five concepts of freedom, and Henry Allison says that this list could "easily be expanded with a little fine-tuning" (Allison, *Kant's*

these concepts have different epistemic statuses. Although reason is unified, it comprises disparate interests (practical and theoretical), and this licenses different justifications for claims about freedom. Thus, the first task will be to define the various concepts of freedom, delineate their boundaries, and explain their connections. Ultimately, transcendental freedom will emerge as the crux of the issue, since Kant claims that this notoriously problematic kind of libertarian freedom is an essential condition of the moral law. It is transcendental freedom that becomes the central focus of the remaining chapters.

Kant is surely to blame for some of these interpretive problems. There is no place in his writings or lectures where he precisely defines and explains of all the different concepts of freedom. On the contrary, the development of his thinking on this matter spanned several decades, and his presentation of the issue varies depending on the task at hand. This has led some commentators to doubt the possibility of presenting a single, coherent explanation of Kant's view. They have not hesitated to express their frustration:

> What is the notion of freedom essential to the *Critique of Pure Reason*? Is it a cosmological freedom, conceived as an idea of reason, and independent of experience? Or is it a practical freedom, known directly through experience? Which of these two types of freedom is related to that freedom which is identical to the autonomous will, and which is required for the *Groundwork of the Metaphysics of Morals*? And what is the relation of intelligible freedom (which according to *Religion within the Limits of Reason Alone*, having already produced evil, converts itself to good) to the autonomous will, which, by definition, cannot remove itself from the moral legislation it imposes — or, on the other hand, to the intelligible freedom of the *Critique of Pure Reason*, which is above all possible change, and hence above every possible conversion? What significance does it have that in the *Critique of Practical Reason* freedom is first deduced as a principle but then admitted as a postulate? It is hardly astonishing, under the circumstances that Kant's theory has been judged obscure and contradictory.[19]

Theory of Freedom, 1). Cf. Lewis White Beck, "Five Concepts of Freedom in Kant." There are a handful of open questions when it comes to tallying up concepts of freedom: should positive and negative conceptions be counted separately? Should we include individual political and legal freedoms? And so on.

[19] Delbos, *La Philosophie Practique du Kant*, 192. Translation from Carnois's *The Coherence of Kant's Doctrine of Freedom*.

These are difficult questions, which cannot be ignored and do not admit of easy answers. My aim in this chapter is to address these issues by presenting a novel, coherent conception of Kant's theory of freedom. I begin by providing a taxonomy of the different kinds of freedom that can be found in Kant's writings. Although commentators have carved it up in a variety of ways, I break it down into four distinct (but interrelated) concepts: practical freedom, transcendental freedom, moral freedom, and political freedom. Each of these categories comprises even more fine-grained concepts, and we must carefully examine them all in order to understand how they form a unified whole. The interpretation I offer is unique, and it has some important advantages over the views that are currently available in the literature. The most noteworthy benefit is that my reading provides a new way of avoiding a popular, uncharitable interpretation known as the "patchwork theory." According to this reading, Kant had not fully worked out a cohesive view of freedom by the time he published the *Critique of Pure Reason*. Instead, proponents of this theory claim that he simply "patched together" his notes from the previous decades. This is supposed to explain the inconsistencies between what Kant says about freedom in the Dialectic and what he writes in the Canon.[20] I explain below what the purported inconsistencies are, and I show how my interpretation undermines the need for such a move.

I conclude the chapter with a brief discussion of the historical development of Kant's view. I clarify the importance of Kant's predecessors, most notably Leibniz, Wolff, Baumgarten, and Crusius, and I then turn to the development of Kant's view from his earliest writings in the 1750s to the critical view that he defends from the 1780s onward.

[20] Among Anglophone scholars, Norman Kemp Smith was one of the first proponents (1918) of the patchwork reading (*A Commentary on Kant's* Critique of Pure Reason, 196). In German scholarship, it appeared almost two decades earlier in Hans Vaihinger's *Die Transzendentale Deduktion der Kategorien*.

1.2 Practical Freedom

Practical freedom is a natural place to start, since the discussion of it inevitably involves both moral and transcendental freedom. Breaking with the interpretive orthodoxy on the taxonomy of Kant's concepts of freedom, I distinguish between two senses of practical freedom. On my interpretation, moral freedom will turn out to be a version of practical freedom, which I call 'strong practical freedom.'[21] It is moral freedom that requires the possibility of transcendental freedom as a necessary condition. I discuss moral and transcendental freedom in the next two sections, but I begin here with a discussion of what I call 'weak practical freedom.'

In many of Kant's discussions of freedom, he differentiates between negative and positive concepts.[22] Practical freedom is no exception. Kant begins with a negative definition of practical freedom, which he describes as "the independence of the power of choice [*Willkür*] from necessitation by impulses of sensibility" (A534 / B562). This is our capacity to resist sensuous impulses or desires, and it is something that we observe on a daily basis:

> Practical freedom can be proved through experience. For it is not merely that which stimulates the senses, i.e., immediately affects them, that determines human choice, but we always have a capacity to overcome impressions on our sensory faculty of desire by representations of that which is useful or injurious even in a more remote way; but these considerations about that which in regard to our whole condition is desirable, i.e., good and useful, depend on reason. (A802–3 / B830–31)

Whenever someone is presented with a desirable object, like a piece of chocolate cake, it is always within the agent's capacity to resist the sensuous impulse for the sake of some other end (e.g., health or future well-being). According to Kant, our possession of this kind of practical freedom is something that separates us from animals:

> For a power of choice [*Willkür*] is **sensible** insofar as it is **pathologically affected** (through moving-causes of sensibility); it is called an **animal** power of choice

[21] To the best of my knowledge, no one in the literature defines the concepts as I do. I argue for the advantages of my reading below in section 1.6.

[22] *Groundwork* III is one example, but this distinction can be found throughout Kant's writings.

> (*arbitrium brutum*) if it can be **pathologically necessitated**. 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)[23]

Like Descartes, Kant thinks that nonhuman animals do not have any capacity for meaningful free agency. He thinks of their actions as completely determined by sensuous impulses; they are basically sophisticated automata.[24] Whenever a desire arises within an animal, it will do as the passion dictates. Humans, by contrast, have the capacity to refrain from sensuous temptation. For Kant, reason is not the slave of the passions, and he claims we know this from everyday experience.

Nevertheless, Kant thinks this definition is incomplete as it describes only the negative sense of practical freedom.[25] Practical freedom has been glossed thus far as freedom *from* necessitation by sensuous desires. In its positive sense, practical freedom is the capacity to act on the motivation of principles given internally by reason rather than those generated externally by the passions. As Kant puts it, "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 from 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

---

[23] It is interesting to note that this discussion of practical freedom in the first *Critique* bears a striking resemblance to his presentation of the issue in his pre-critical writings and lectures. For example, the following is an excerpt from one of Kant's metaphysics lectures in the 1770s: "Accordingly, the human power of choice <*arbitrium humanum*> is not brute <*brutum*>, but rather free <*liberum*>. This is the power of free choice <*arbitrium liberum*>, so far as it is defined psychologically or practically. However, that power of choice <*arbitrium*> which is not necessitated or impelled at all by any stimuli <*stimulos* >, but rather is determined by motives, by motive grounds of the understanding, is the intellectual or transcendental power of free choice <*liberum arbitrium*>. The sensitive power of choice <*arbitrium sensitivum*> can indeed be free <*liberum*>, but not the brute one <*brutum*>. The sensitive power of [free] choice <*arbitrium sensitivum [liberum]*> is only affected or impelled by the stimuli <*stimulis*>, but the brute one <*brutum*> is necessitated." (25:255)

[24] In his lectures he bluntly states "Animals can be necessitated strictly through stimuli <*stricte per stimulos*>" (L_1 28:256). Cf. the Mrongovius lectures: "Our prerogative is that we are free from the laws of sensibility. Animals are not, and are therefore merely passive" (*MMr* 29: 902)

[25] The idea that negative definitions require positive counterparts is something of a refrain in Kant's writings and lectures: "But we still must have a positive concept of it [freedom], otherwise we have no concept at all" (*MMr* 29: 902).

part of pure and, as such, practical reason is freedom in the *positive* sense" (*KpV* 5:33). Practical freedom in the positive sense is our capacity to act according to principles that reason gives itself. This, of course, requires freedom in the negative sense: for if actions were determined entirely by the passions, then reason would be impotent.

This capacity for restraint does not, however, demonstrate that human beings have libertarian free will. If practical freedom is defined in this way (i.e. acting on behalf of an end given by reason rather than one given externally by the passions) then it would be something that a hard determinist could endorse.[26] Our knowledge that we have practical freedom does not license us to claim that a person could have acted differently than she did in the actual world. An action which shows restraint from sensuous temptation might simply have been determined by other, less apparent (but equally deterministic) causes. That action might have been necessitated. But Kant is aware of this point.[27] In order for him to claim that an agent could have acted differently than she did in the actual world, something other than practical freedom is required. In its negative aspect, 'weak practical freedom' simply means that choices are not necessitated by sensible stimuli (passions, desires, inclinations, etc.). In the positive sense, it is our capacity to act according to act on prescriptions given by reason.

This empirically knowable, compatibilist concept of freedom is not sufficient for Kant's purposes. His concern with morality commits him (for better or worse) to a libertarian concept. In the first *Critique*, he considers the case of a person who has told a malicious lie, and he puts a great

[26] Again, this interpretation of mine is controversial. It is rejected, for instance, by Allison who understands practical freedom as an incompatibilist concept. See Allison, *Kant's Theory of Freedom*, 54. In my discussion of the patchwork theory, I explain why we should reject Allison's suggestion that practical freedom requires (or is identical with) libertarian freedom.

[27] See, for example, *MMr* 29:879.

deal of weight on the freedom to do otherwise as a necessary condition of legitimately attributing blame:

> [O]ne may take a voluntary action, e.g. a malicious lie, through which a person has brought about a certain confusion in society; and one may first investigate its moving causes, through which it arose, judging on that basis how the lie and its consequences could be imputed to the person. With this first intent one goes into the sources of the person's empirical character, seeking them in a bad upbringing, bad company, and also finding them in the wickedness of a natural temper insensitive to shame, partly in carelessness and thoughtlessness; in so doing one does not leave out of account the occasioning causes. In all this one proceeds as with any investigation in the series of determining causes for a given natural effect. Now *even if one believes the action to be determined by these causes, one nonetheless blames the agent*, and not on account of his unhappy natural temper, not on account of the circumstances influencing him, not even on account of the life he has led previously; *for one presupposes that it can be entirely set aside how that life was constituted, and that the series of conditions that transpired might not have been, but rather that this deed could be regarded as entirely unconditioned in regard to the previous state*, *as though with that act the agent had started a series of consequences entirely from himself*. This blame is grounded on the law of reason, *which regards reason as a cause that, regardless of all the empirical conditions just named, could have and ought to have determined the conduct of the person to be other than it is*. (A554–55 / B582–83; emphasis added)

It is evident from what Kant says here that something other than practical freedom is required. He makes it clear that the legitimacy of our imputation of blame is grounded in the fact that the agent "could have and ought to have" done otherwise. And, as we have seen, the concept of weak practical freedom, as I have described it here, does not license us to make such a claim. Even Spinoza could agree that we have this kind of practical freedom, and he firmly rejects the idea that an agent could have done otherwise.[28] If Kant is going to defend the claim that "this deed could be regarded as entirely unconditioned in regard to the previous state, as though with that act the agent

[28] Spinoza emphasizes this point throughout the *Ethics*. According to Spinoza, we have no free will, and things could not have turned out differently: "Things could not have been produced by God in any way or in any order other than they were (Ip33). "In the Mind there is no absolute or free will, but the Mind is determined to willing this or that by a cause, which was also determined by another, and this again by another, and so on to infinity" (IIp48; translations mine). But Spinoza's rejection of libertarian free will does not prevent him from accepting the possibility of weak practical freedom. Indeed, one of the principal aims of the *Ethics* is to show how human beings can (and should) resist the impulses of the passions and act according to the dictates of reason.

had started a series of consequences entirely from himself" then he is going to need something other than weak practical freedom.

1.3 Transcendental Freedom

Since Kant finds practical freedom insufficient for the task at hand, he turns to another concept. Unfortunately, the kind of freedom which satisfies his constraints on imputing praise and blame is the very concept that entangles reason in contradictions:

> The transcendental idea of freedom . . . constitutes only that of the absolute spontaneity of an action, as the real ground of its imputability; but this idea is nevertheless the real stumbling block for philosophy, which finds insuperable difficulties in admitting this kind of unconditioned causality. (A448 / B476)

He concludes that transcendental freedom alone is capable of grounding our moral concepts of praise, blame, autonomy, etc. His moral philosophy will therefore require libertarian freedom. But this does not mean moral considerations constitute the sole impetus for transcendental freedom. In the first *Critique*, Kant claims that it is a necessary concept of theoretical reason.[29] It is what he calls a "transcendental idea" — one which is indispensable even though cannot be known through experience and cannot be a condition of the possibility of experience:[30]

> By freedom in the cosmological sense, on the contrary, I understand the faculty of beginning a state from itself, the causality of which does not in turn stand under another cause determining it in time in accordance with the law of nature. Freedom in this signification is a pure transcendental idea, which, first, *contains nothing*

[29] The role of transcendental freedom in Kant's practical philosophy is the focus of chapter 3, and I discuss its place in his theoretical philosophy in chapter 2.

[30] This will be explained in much greater detail in the next chapter. The important point is that the pure transcendental ideas of reason can never be encountered in experience: "By the idea of a necessary concept of reason, I understand one to which no congruent object can be given in the senses. Thus the pure concepts of reason we have just examined are transcendental ideas. They are concepts of pure reason; for they consider all experiential cognition as determined through an absolute totality of conditions. They are not arbitrarily invented, but given as problems by the nature of reason itself, and hence they relate necessarily to the entire use of the understanding. Finally, they are transcendent concepts, and exceed the bounds of all experience, in which no object adequate to the transcendental idea can ever occur" (A327 / B383–84).

> *borrowed from experience, and second, the object of which also cannot be given determinately in any experience*. (A533 / B561)

But even though the concept has no empirical content, Kant has a determinate idea of what transcendental freedom consists in. The definitions of it that are scattered throughout his writings and lectures are strikingly similar:[31]

> *Critique of Pure Reason*
> Accordingly, a causality must be assumed through which something happens without its cause being further determined by another previous cause, i.e., an absolute causal spontaneity beginning from itself [*von selbst*] a series of appearances that runs according to natural laws, hence transcendental freedom . . .
>
> **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 in accordance with constant laws. (A445–46 / B473–74)
>
> *Critique of Practical Reason*
> But if no determining ground of the will other than that 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. But such independence is called *freedom* in the strictest, that is, in the transcendental, sense. (5:29)
>
> [*T*]*ranscendental freedom,* which must be thought as independence from everything empirical and so from nature generally, whether it is regarded as an object of inner sense in time only or also of outer sense in both space and time; without this freedom (in the latter and proper sense), which alone is practical a priori, no moral law is possible and no imputation in accordance with it. (5:97)
>
> Metaphysics Lectures
> That freedom, however, which is wholly independent of all stimuli <*stimulis*>, is transcendental freedom . . . (*L1* 28:257 [Mid-1770s])
>
> The soul is a being which acts spontaneously, simply speaking <*simpliciter spontan*>; i.e., the human soul is free in the transcendental sense <*in sensu*

[31] This holds from the 1770s onward. In his earliest writings such as the *New Elucidation*, he endorses a Leibnizian/Wolffian concept of freedom, and, at that time, his commitment to the principle of sufficient reason rules out the possibility that an agent could have done otherwise.

> *transcendentali*> . . . this means absolute spontaneity, and is self-activity from an *inner principle* according to the power of free choice. (*L_1* 28:267–78 [Mid-1770s])
>
> Freedom is the faculty for beginning a series of states oneself. (*MMr* 29:861 [1782–83)])

From these definitions, we can identify the central features of transcendental freedom. Once again, it is useful to isolate the negative and positive aspects.

In the negative sense, transcendental freedom requires complete independence from the empirical world and the laws of nature. In order to be transcendentally free, the agent must not be causally necessitated by any antecedent states (even internal ones). And since all events in space and time are causally determined by the laws of nature, transcendental freedom requires an "intelligible cause," which is outside of time (A537 / B565). This means that transcendental freedom can be predicated only of noumenal beings: "[I]f one still wants to save it [freedom], no other path remains than to ascribe the existence of a thing so far as it is determinable in time, and so too its causality in accordance with the law of natural necessity, only to appearance, *and to ascribe freedom to the same being as a thing in itself*." (*KpV* 5:95; emphasis added). In sum: transcendental freedom requires independence from the laws of nature, sensible necessitation, time, and the empirical world.

And in addition to these negative conditions, transcendental freedom has a positive definition. Most notably, a transcendentally free agent must be capable of acting "spontaneously" or "from itself" (*von selbst*).[32] But 'spontaneity' here has a special meaning. In the two thousand years between Plato and Kant, the concept of spontaneity had taken a variety of forms,[33] and it was

[32] '*Von selbst*' literally translates to 'by itself' or 'from itself,' but it is often glossed as 'spontaneous.' Kant also uses the Latin '*spontaneitas'* and its German cognate '*Spontaneität*.'

[33] According to Marco Sgarbi, the earliest known mention of 'spontaneity' [*hekousious* in Greek] is found in a fragment from Democritus, and he uses the term to refer to "simple actions that do not arise from reflection" (Sgarbi,

especially prominent in the writings of Leibniz and Wolff. Kant is at pains to distinguish his concept of spontaneity from theirs, since he came to reject their view in the years leading up the publication of the first *Critique*. On Leibniz's view spontaneity does not require the ability to do something other than what was done in the actual world. Such libertarian freedom would contravene his commitment to the principle of sufficient reason. For Leibniz, spontaneity consists in *not* being coerced or subject to external causal necessitation. And, because of his doctrine of pre-established harmony (according to which every state of a substance is causally determined internally by that substance alone), spontaneity follows immediately from Leibniz's metaphysical commitments:

> [T]he establishment of this system [of preestablished harmony] demonstrates beyond a doubt that in the course of nature each substance is the sole cause of all its actions, and that it is free of all physical influences from every other substance, save the customary cooperation of God. And this system shows that our spontaneity is real, and not only apparent . . . The spontaneity of our actions can therefore no longer be questioned; and Aristotle has defined it well, saying that an action is spontaneous when its source is in him who acts . . . Thus it is that our actions and our will depend entirely upon us. (G 6:295–96 / *Theodicy* 300–301)

Kant initially accepted Leibniz's concept of spontaneity and even defended it against critics like Crusius. But by the 1770s, he had substantially revised his position. In his lectures from that decade, he distinguishes between absolute spontaneity (*absoluta vel simpliciter talis*) and spontaneity which is qualified in some respect (*secundum quid talis*). He thinks that qualified spontaneity is the kind we see in a machine or automaton as long as it "moves itself according to an inner principle, e.g., a watch, a turnspit" (L_1 28:267). In the *Critique of Practical Reason*, he repeats the claim about the turnspit, and he makes it clear that Leibniz is his target:

Kant on Spontaneity, ch. 2). Plato mentions it in the *Cratylus* and the *Sophist*; Aristotle uses it in the *Nicomachean Ethics* where he distinguishes voluntary actions from involuntary ones. And it is subsequently picked up by Scholastics and Protestants alike: Calvinists, Lutherans, Jesuits, etc. As these debates went on, the range of meanings for 'spontaneity' widened considerably. The dominant paradigm in Kant's historical context, however, came from Leibniz.

> Here one looks only to the necessity of the connection of events in a time series as it develops in accordance with natural law, whether the subject in which this development takes place is called *automaton materiale,* when the machinery is driven by matter, or with Leibniz *spirituale,* when it is driven by representations; and if the freedom of our will were none other than the latter (say, psychological and comparative but not also transcendental, i.e., absolute), then it would at bottom be nothing better than the freedom of a turnspit, which, when once it is wound up, also accomplishes its movements of itself. (*KpV* 5:97)

Kant thinks that the Leibnizian concept of spontaneity is inadequate and that only transcendental freedom will suffice. The concept of spontaneity that Kant has in mind must be something other than being determined by internal states. Indeed, it seems that what Kant requires is a first cause — one that is not in any way determined by antecedent conditions. He calls it "a first beginning of a series of appearances" (A449 / B477): a capacity to "begin a series of occurrences entirely from itself" (A534 / B 562). This kind of spontaneous action requires that "nothing precedes it through which this occurring action is determined in accordance with constant laws" (A445 / B473). This is what Kant means by 'spontaneity' when it comes to transcendental freedom.

To avoid confusion, it should be noted that the word 'spontaneity' does double duty in the first *Critique.* In addition to this role that it plays in the context of transcendental freedom, spontaneity is also prominently featured in Kant's discussion of the understanding. He distinguishes the passive receptivity of the sensibility with the active spontaneity of the understanding. Although the sensibility passively receives external stimuli, the understanding actively synthesizes representations into a manifold by means of concepts: "If we will call the receptivity of our mind to receive representations insofar as it is affected in some way sensibility, then on the contrary the faculty for bringing forth representations itself, or the spontaneity of cognition, is the understanding" (A51 / B75). There is nearly universal agreement among commentators that the spontaneity of understanding is distinct from transcendental freedom's

ability to spontaneously begin a series of appearances.[34] The spontaneity of the understanding is merely relative; it is spontaneity *secundum quid talis*. It is spontaneous only in the sense that it is directed by an inner principle rather than an external one. The spontaneity of transcendental freedom, however, is absolute spontaneity. It is the capacity to act as a first cause of a series of appearances.

And now we have a complete definition of transcendental freedom. Negatively construed, it is total independence from laws of nature, sensible necessitation, time, and the empirical world, and in its positive sense it is the capacity to spontaneously beget a series of appearances.

1.4 Moral Freedom

As I mentioned above, Kant famously asserts a relationship of identity between freedom and morality. In the *Groundwork* he says that "a free will and a will under moral laws are one and the same" (*G* 4:447). And again in the second *Critique* he writes that "the moral law expresses nothing other than the *autonomy* of pure practical reason, that is, freedom" (5:33). In chapter 3, I look closely at Kant's argument that freedom and morality "reciprocally imply each other," but first we must understand the concepts that are being identified. What kind of freedom is at stake here? The concept in this domain is autonomy. And in keeping with its etymology,[35] Kant uses this term to denote the capacity for self-legislation — the ability to act according to a law that reason gives itself: "Autonomy of the will is the property of the will by which it is a law to itself

[34] Sellars argued for this view in 1970 and few have disagreed with his assessment. Markus Kohl is one of the only dissenters. He argues that the spontaneity of our empirical judgments is essentially the same as the spontaneity of practical reason. See Kohl, "Kant on Freedom of Empirical Thought." Kohl is aware that he is virtually alone on this: "Wilfrid Sellars famously suggests that Kant uses the absolute conception of spontaneity in his account of free moral agency but invokes the merely relative conception when he refers to the spontaneity of understanding in his epistemology, a point that has been widely (albeit not universally) accepted by commentators." See Kohl's review of *Kant on Spontaneity*.

[35] From Attic Greek, '*auto*' means 'self,' and '*nomos*' means 'law.'

(independently of any property of the objects of volition)." (*G* 4:440). Kant considers this to be "supreme principle of morality," and it is contrasted with heteronomy:

> 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 that case does not give itself the law; instead the object, by means of its relation to the will, gives the law to it. This relation, whether it rests upon inclination or upon representations of reason, lets only hypothetical imperatives become possible: I ought to do something *because I will something else.* (*G* 4:441)[36]

Whenever we act according to the law that reason gives itself (namely the moral law), we act autonomously, and whenever we act for the sake of some externally given inclination (e.g., pride, honor, wealth, health, etc.) we act heteronomously.

Kant then identifies autonomy with freedom of the will: "Natural necessity was a heteronomy of efficient causes, since every effect was possible only in accordance with the law that something else determines the efficient cause to causality; **what, then, can freedom of the will be other than autonomy, that is, the will's property of being a law to itself**?" (*G* 4:446–47; emphasis added). Once again, Kant separates the negative and positive components. Autonomy requires that we not be "determined to activity by the influence of alien causes" (*G* 4:446), and it consists in reason's capacity for self-legislation.

---

[36] Cf. Kant's Ethics lectures: "Now if moral actions are to be grounded in the form of lawfulness, the moral laws must have their basic determination in a law-giving power which constitutes legislation. Moral legislation is the law-giving of human reason, as which it is the law-giver in regard to all laws, and is so through itself. This is the autonomy of reason, whereby, that is, it determines the laws of free choice through its own law-giving, independently of any influence, and the principle of the autonomy of reason is thus the individual legislation of choice by reason. The opposite would be heteronomy, i.e., legislation that is founded in like fashion on grounds other than the freedom of reason.

Thus if, for example, the principle of universal happiness were to be the basis for determination of the moral laws, it would be a question of how far our needs were satisfied in their entire totality by following these laws; but here laws of nature are involved, and the moral laws would have to be subject to them, so that reason would have to obey the laws of nature and sensibility, and that in a necessary fashion (for in the physical order this is so anyway). But this would obviously put an end to the autonomy of reason, and thus be heteronomy." (27: 499)

At first glance, it might seem like autonomy is nothing other than practical freedom. After all, practical freedom was defined as (1) independence from necessitation by sensuous impulses and (2) the capacity to act on the basis of reason. That seems dangerously close to the above definition of autonomy. But these two concepts must be distinguished from one another if Kant is to avoid inconsistency. Practical freedom is wider than autonomy, for it encompasses actions which must be described as heteronomous. The distinction is fairly clear if we contrast the definition of practical freedom from the Canon with the conditions of autonomy. Again, on my view, weak practical freedom is proved through experience whenever we overcome sensuous stimuli and act out of considerations from reason including (but not limited to) what is desirable, good, useful, healthy, etc. If I am tempted by sensuous impulses to waste money on a luxury good but I refrain from doing so in order to preserve my future wealth, then I have exercised my practical freedom. I have not necessarily acted autonomously. On the contrary, I have made an object (my future prosperity) the determining ground of my will. This sort of imperative can only be hypothetical: "I ought to refrain from wasteful spending, because I want to be as wealthy as possible when I am older." Moral imperatives must be categorical and hold for all rational beings; whereas hypothetical imperatives hold only on the condition that some contingent desire is present:

> Wherever an object of the will has to be laid down as the basis for prescribing the rule that determines the will, there the rule is none other than heteronomy; the imperative is conditional, namely: if or because one wills this object, one ought to act in such or such a way; hence it can never command morally, that is, categorically . . . *I ought to do something on this account, that I will something else*, and here yet another law must be put as a basis in me, the subject, in accordance with which I necessarily will this something else, which law in turn needs an imperative that would limit this maxim . . . it would, strictly speaking, be nature that gives the law; and this, as a law of nature, must not only be cognized and proved by experience — and is therefore in itself contingent and hence unfit for an apodictic practical rule, such as moral rules must be — but it is *always only heteronomy* of the will; the will would not give itself the law but a foreign impulse would give the law to it by means of the subject's nature, which is attuned to be receptive to it. (*G* 4:444)

Practical reason generates a variety of precepts which are merely hypothetical imperatives. For example, Kant says that an important practical precept is that a young man "must work and save in his youth in order to not want in his old age," but he notes that such a rule commands compliance only if the end (not wanting in old age) is desirable (*KpV* 5:20). The rule, as he points out, would mean little to someone who expects to die young or does not care about material possessions.

Although practical reason produces many such precepts which involve the exercise of practical freedom, only a proper subset of these require autonomy. Practical reason can ask us to act against our sensuous desires for the sake of many ends (health, honor, prosperity, happiness, etc.), but autonomy "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 maxim must be capable of" (*KpV* 5:33). Thus, autonomy is a special kind of practical freedom, which I have called "strong practical freedom." We exercise this kind of freedom and act autonomously only if we act from the motive of duty. All practical freedom involves acting on behalf of reason, but only *some* of those actions come from *pure* practical reason.

1.5 Political Freedom

Last but not least, there is political freedom. On Kant's view, there is only one innate right: freedom. In the *Metaphysics of Morals* he writes, "*Freedom* (independence from being constrained by another's choice), insofar as it can coexist with the freedom of every other in accordance with a universal law, is the only original right belonging to every man by virtue of his humanity"

(6:237).[37] As the only innate right, freedom is the only legitimate basis of the state. Kant rejects every other foundation, especially welfare or happiness. His argument for this claim mirrors the one he gives in the *Groundwork* and the second *Critique* as to why happiness cannot ground morality: happiness is too variable and contingent. We all have our own conception of happiness, and the state is in no position to impose any particular conception of the good life on anyone else, for that would be a limitation on our freedom.[38] Indeed, to do so would be to engage in a kind of unacceptable paternalism. Kant thinks that "no one can coerce me to be happy in his way" and that a state which does this violates its subjects' freedom and thus constitutes the "greatest despotism thinkable" (*TP* 8:290–91).

Kant's political philosophy makes an important contribution to the social contract tradition, and his presentation of it is notably distinct from Hobbes's.[39] Hobbes conceives of the state of nature as one of maximal liberty. Because there is no state with coercive power, everyone is free to satisfy needs and desires by any means necessary. As Hobbes puts it, "It followeth, that in such a condition, every man has a Right to every thing; even to one another's body."[40] People choose to give up some of the liberties they enjoy in the state of nature because it is a "war of all against all" in which life is "solitary, poor, nasty, brutish, and short."[41] So conceived, the coercive power

[37] This is derived from Kant's commitment to what he calls the universal principle of right: "[A]n action is *right* if it can coexist with everyone's freedom in accordance with a universal law, or if on its maxim the freedom of choice of each can coexist with everyone's freedom in accordance with universal law" (*MS* 6:230)

[38] This is quite similar to John Rawls's idea in *Political Liberalism* regarding legitimacy. Since everyone has their own idea of the good (what Rawls calls a "comprehensive doctrine"), no law can be legitimate in a liberal state unless it can expect assent among reasonable people with different concepts of the good. "Since justification is addressed to others, it proceeds from what is, or can be, held in common; and so we begin from shared fundamental ideas implicit in the public political culture in the hope of developing from them a political conception that can gain free and reasoned agreement in judgment." (*Political Liberalism*, 100–101)

[39] Unsurprisingly, Kant's view is closer to Rousseau's in *On the Social Contract*.

[40] *Leviathan* Part I, ch. 14, section 4.

[41] *Leviathan*, Part I, ch. 8, section 9.

of the state is a limitation of freedom but one which is justified because the sovereign guarantees the basic security of its subjects and this safety is preferable to the condition of war.

Kant sees it quite differently. We should leave the state of nature *not* simply out of a prudential concern for our safety; we are *morally* required to leave the state of nature and establish a "rightful condition" of law. Kant acknowledges that no one violates any legal rights in the state of nature (since there can be no such rights), but they nevertheless "do wrong in the highest degree by willing to be and to remain in a condition that is not rightful, that is, in which no one is assured of what is his against violence" (*MS* 6:308–09). And thus he concludes that since "you cannot avoid living side by side with all others, you ought to leave the state of nature and proceed with them into a rightful condition [*rechtliche Zustand*]" (*MS* 6:307).[42]

Under the Hobbesian conception of the social contract, the creation of the state legitimates coercion and can be thought of as a constraint on freedom, albeit one that is prudent to accept. Once again, Kant disagrees. Although he concedes that "any limitation of freedom through another's choice is called coercion" and that a "civil constitution is a relation of *free* human beings who . . . are nevertheless subject to coercive laws," he thinks that the rightful condition of law is not only consistent with but indeed a necessary condition of maximal freedom (*TP* 8:290). This is because we are to think of the state's coercive acts not simply as a hindrance to freedom. To do so would be to think of the state's power as a violation of our only innate right. We should instead think of legitimate coercion as the "hindering of a hindrance of freedom" (*MS* 6:231).[43] The state

[42] Of course, it is not just violence or safety that motivates this move. One of Kant's principal considerations is the possibility of property rights. In the state of nature, no one is secure in their property. They can only maintain the properties that they can physically retain and defend. The transition in the *Metaphysics of Morals* from private right to public right largely focuses on Kant's discussion of property. Again, Kant's view of the social contract borrows more heavily from Rousseau than it does from Hobbes.

[43] In *Force and Freedom*, Arthur Ripstein provides a careful and comprehensive articulation of Kant's political philosophy. As he puts it, "Coercion is objectionable where it is a hindrance to a person's right to freedom, but legitimate when it takes the form of hindering a hindrance to freedom. To stop you from interfering with another person upholds the other's freedom" (55).

can legitimately use its power to forcibly prevent a murderer from killing. Rather than thinking of this coercion as a limitation of the murderer's freedom, we recognize how the murderer's actions infringe on the victim's freedom and thus see the coercion as protecting rather than limiting freedom.

Freedom is therefore playing several important roles in Kant's political philosophy. (1) It is the only innate right. (2) It is the sole justification for the existence of the state. And (3) it is the condition of legitimacy for the state's laws. There are other manifestations of political freedom in Kant's thought, but these are largely applied considerations that are downstream of the more basic commitments. That is not to say, however, that these applied forms of freedom are insignificant; many of them were very important to Kant. For instance, he was a champion of the Enlightenment, and he argues that the freedom "to make public use of one's reason in all matters" is the only necessary condition of enlightenment.[44] Similarly, he argues that "freedom of the pen" is the "sole palladium of the people's rights" (*TP* 8:304). Although these applied forms of political freedom follow from the above commitments, they are not for that reason any less significant.

The final consideration of political freedom concerns what lies upstream of Kant's political commitments. From the outset it was taken as a given that freedom is the only innate right, and the remainder of his political philosophy was to be derived from this idea. Kant does not expect us to accept this as a bare assertion or brute fact; rather, it should be understood as a consequence of his moral philosophy. And without understanding the connection between the two, we would fail to appreciate the unity of Kant's moral and political concepts of freedom.

[44] Enlightenment is construed as "the human being's emergence from his self-incurred minority [*Unmündigkeit*]." That is, people should have the courage to think for themselves. And Kant says, "For this enlightenment, however, nothing is required but *freedom* . . . to make public use of one's reason" (*E* 8:37).

Recognizing freedom as an innate political right is something that is supposed to follow from Kant's claim that rational beings are ends in themselves and that their capacity for rational agency (and thus morality and freedom) confers upon them an unconditional, non-fungible worth which Kant calls "dignity" (*Würde*). "In the kingdom of ends everything has either a price or a dignity" (*G* 4:434). All items in the marketplace have a price and can be exchanged for something of equivalent value. Human beings, by contrast, are regarded as ends in themselves and thus their value is not exchangeable. But this only pushes the inquiry back a step. Why should we elevate the value of rational nature above everything else? Kant makes it clear that freedom is the ground of this exceptional worth:

> And what is it, then, that justifies a morally good disposition, or virtue, in making such high claims? It is nothing less than the *share* it affords a rational being *in the giving of universal laws*, by which it makes him fit to be a member of a possible kingdom of ends, which he was already destined to be by his own nature as an end in itself and, for that very reason, as lawgiving in the kingdom of ends — **as free with respect to all laws of nature, obeying only those which he himself gives and in accordance with which his maxims can belong to a giving of universal law** (to which at the same time he subjects himself). For, nothing can have a worth other than that which the law determines for it. **But the lawgiving itself, which determines all worth, must for that very reason have a dignity, that is, an unconditional, incomparable worth**; and the word *respect* alone provides a becoming expression for the estimate of it that a rational being must give. *Autonomy* is therefore the ground of the dignity of human nature and of every rational nature. (*G* 4:435–36; boldface added).

This brings the account full circle. Kant's political philosophy is premised on the idea that freedom is the only innate right. This is grounded on the moral principle that freedom (more specifically, autonomy) is the only possible basis for unconditional value. And autonomy was defined as a special kind of practical freedom. The metaphysical underpinning of autonomy is transcendental freedom. This shows just how much Kant has riding on the concept of freedom — transcendental freedom especially.

1.6 The Patchwork Theory

Unfortunately, transcendental freedom's connection to practical freedom turns out to be somewhat problematic. As I mentioned above (and I explain in greater detail in the next chapter), theoretical reason must remain agnostic about transcendental freedom. There is nothing from which speculative reason can infer that human beings have transcendental freedom. But transcendental freedom was also described as a necessary condition of practical freedom. Kant says that transcendental freedom is the ground of practical freedom and the latter would be destroyed if the former were abolished.[45] This seems to be inconsistent with Kant's claim that practical freedom is empirically demonstrable. It would appear that Kant has committed himself to an inconsistent triad:

> (1) Transcendental freedom is a necessary condition of practical freedom.
> (2) We know from experience that we have practical freedom.
> (3) We cannot know and must be agnostic about our possession of transcendental freedom.

As I said in the introduction, many commentators have dealt with this problem by conceding that Kant is inconsistent on this matter. They explain the mistake by claiming that Kant did not have a coherent view worked out by 1781; instead, they suggest that he patched together some of his notes from previous decades.[46]

---

[45] A533–34 / B561–62

[46] Norman Kemp Smith, Hans Vaihinger, and Erich Adickes were the earliest proponents of the patchwork theory, and H.J. Paton was one of the first detractors. See Paton, *Kant's Metaphysics of Experience*, ch. 16. The patchwork theory is still fairly popular, as it is an easy way to make sense of the inconsistency. For support of the patchwork theory, see Carnois *The Coherence of Kant's Doctrine of Freedom*; Martial Guéroult: "Canon de la Raison pure et *Critique de la Raison Pratique*"; and Dieter Schönecker, *Kants Begriff transzendentaler und praktischer Freiheit. Eine entwicklungsgeschichtliche Studie*.

Although the patchwork theory is still quite popular, several commentators have suggested ways of avoiding it. See, for example, Allison, *Kant's Theory of Freedom*, ch. 3; Julio Esteves, "The Alleged Incompatibility between the Concepts of Practical Freedom in the Dialectic and in the Canon of the *Critique of Pure Reason*"; and Markus Kohl, "Transcendental and Practical Freedom in the *Critique of Pure Reason*."

My interpretation opens up a new way of solving the interpretive problem. By distinguishing between strong and weak forms of practical freedom, the triad is no longer inconsistent:

> (1) Transcendental freedom is a necessary condition of **strong** practical freedom.
> (2) We know from experience that we have **weak** practical freedom.
> (3) We cannot know and must be agnostic about our possession of transcendental freedom.

On my view transcendental freedom is not a necessary condition of every instance of practical freedom; it is required only for strong practical freedom (i.e., moral freedom/autonomy). Although experience attests to our ability to exercise weak practical freedom (e.g., abstaining from eating the chocolate cake), we cannot know from experience that we have acted from duty. Indeed, in the *Groundwork* Kant concedes that experience cannot tell us whether anyone in history has ever acted from the pure motive of duty.[47] Thus, we do not have any empirical justification for the claim that human beings have transcendental freedom, and this is how Kant thinks it should be.

I believe my reading has some advantages over those in the literature. First, the patchwork theory is a rather uncharitable interpretation, and I think we should avoid attributing such a glaring inconsistency to Kant on this matter unless we have no way of avoiding it. On the other side, there have been some attempts to resist the patchwork reading by denying that the first *Critique* contains any empirical, compatibilist concept of practical freedom.[48] Instead, these commentators understand practical freedom to be a strictly incompatibilist form of freedom. The problem with such a reading is that it faces the difficult task of explaining the sections from Canon in which Kant seems to say rather explicitly that experience proves that we have practical freedom. The

[47] "[E]ven if there never have been actions arising from such pure sources, what is at issue here is not whether this or that happened; that, instead, reason by itself and independently of all appearances commands what ought to happen . . ." (G 4:407–8).

[48] See, for example, Kohl, "Transcendental and Practical Freedom in the *Critique of Pure Reason*" and Allison, *Kant's Theory of Freedom*.

natural reading of such passages would commit them to the conclusion that experience proves the reality of libertarian freedom: something that Kant would surely deny. In response, they have argued that the Canon is operating with a "non-standard," "aberrant" sense of empirical proof, and that this allows them to avoid the unpalatable conclusion that experience proves the reality of libertarian freedom.[49] Although I accept the possibility of such an interpretation, I believe that mine is preferable insofar as it seems to be the most straightforward way of rendering Kant's claim that "Practical freedom can be proved from experience" (A802 / B830).[50]

If we read the first *Critique* in the way that I am suggesting, then we can see that it contains both compatibilist and incompatibilist concepts of freedom. This is not terribly surprising if we take into account how Kant's view of freedom changed over time. For years, he worked only with a compatibilist concept of freedom, and it was not until the decades leading up to the first *Critique* that he came to think of libertarian free will as having a legitimate place in his philosophical system. But he never dispensed with the compatibilist concept entirely. In order to better understand the nature of Kant's shift, I discuss the both the historical background and the development of Kant's theory of freedom in the next section.

1.7 The Historical Development of Kant's Concept of Freedom

Kant's stance on freedom was heavily influenced by the German rationalist tradition, and in many cases his positions were taken directly from Leibniz, Wolff, Baumgarten, and Crusius. In other instances, especially the critical case, his view arose out of his dissatisfaction with the views

[49] Kohl, "Transcendental and Practical," 323.

[50] "Die praktische Freiheit kann durch Erfahrung bewiesen werden." I think this weighs in my favor, but I do not take it to be a decisive consideration in my favor. It is not my intention to completely rule out other suggestions of ways to avoid the patchwork theory. I simply want to highlight how my interpretation of practical freedom opens up a new possibility.

on offer. Kant derived his pre-critical view of freedom from two sources: (1) Leibniz and those who defended his views of freedom and the principle of sufficient reason and (2) Crusius, who rejected Leibniz's compatibilism. Kant received much of his Leibnizian philosophy second-hand from one of its most influential proponents, Christian Wolff. Wolff's writings on theology, psychology, logic, and ontology had a lasting impact on Kant. And the same could be said of Baumgarten, whose *Metaphysica* served as the textbook for Kant's lectures on metaphysics and thus provided the conceptual framework that Kant used to articulate his views throughout his career.[51]

When it comes to the historical context of Kant's concept of freedom, Leibniz's work is the most natural place to begin. As noted above, Leibniz considered spontaneity to be one of the necessary conditions of freedom, but it was not the only one. In his *Theodicy*, Leibniz writes, "I have shown that freedom, according to the definition required in the schools of theology, consists in intelligence, which involves a clear knowledge of the object of deliberation, in spontaneity, whereby we determine, and in contingency, that is, in the exclusion of logical or metaphysical necessity" (§288). In addition to spontaneity, Leibniz includes intelligence and contingency as conditions of freedom. Intelligence is a property of rational beings (i.e., those endowed with souls), and it comprises the capacity to judge something as good and thus "inclines the will without necessitating it" (*ibid*). Although various scholastic thinkers included "indifference" as a condition of freedom,[52] early modern thinkers like Descartes and Leibniz cast it a negative light. Descartes describes indifference as "the lowest grade of freedom" saying that it "attests not to any perfection

[51] Kant, *Lectures on Metaphysics,* xiv.

[52] For example, William of Ockham and Peter John Olivi were proponents of a libertarian view of free will that requires indifference.

of freedom, but rather to a defect in knowledge or a kind of negation." (AT 7:57–58/ CSM 2:40).[53] Leibniz similarly derided the idea of being indifferent between alternative possibilities; he not only considered "moral necessity" compatible with freedom, he upheld it as a blessed kind of necessity — the sort that compels God to freely choose the best possible world. And in his earliest metaphysical writings, Kant was willing to endorse this position as well. Even as late as the 1780s, Kant talks about freedom in terms of degrees saying that freedom is directly proportional to the extent to which rationality compels us to do what the moral law requires: "The more a man can be morally compelled the freer he is; the more he is pathologically compelled, though this only occurs in a comparative sense, the less free he is" (*Collins* 27:268). Thus, following Leibniz, Wolff and Baumgarten, Kant takes intelligence — understood as the capacity to judge what it is best and to act in accordance with this judgment — to be a necessary condition of freedom, especially moral freedom or autonomy.

Contingency was a different matter: one that Kant changes his mind on after decades of reflection. According to Leibniz, a human action, such as Caesar crossing the Rubicon, is contingent in the sense that it is logically possible for the agent to do otherwise.[54] Leibniz thought that this sense of logical contingency was compatible with the action being necessitated by deterministic laws and the antecedent conditions of Caesar's soul. Being the person he is,[55] there was no way Caesar could have chosen not to cross the Rubicon, but Leibniz thinks that this is merely "hypothetical necessity." It does not mean that Caesar's action was not free, for it is logically possible that God could have created a different being, Caesar*, who did not cross the

[53] Cf. Descartes to Mesland 1645: "As for freedom of the will . . . I would like you to notice that "indifference" in this context seems to me strictly to mean that state of the will when it is not impelled one way rather than another by any perception of truth or goodness. This is the sense in which I took it when I said that the lowest degree of freedom is that by which we determine ourselves to things to which we are indifferent" (AT 4:173/ CSMK 244–45).

[54] See, for example, *Discourse on Metaphysics* §13.

[55] On Leibniz's view, the complete concept of Caesar contains the predicate "crosses the Rubicon" and thus the claim "Caesar crosses the Rubicon" is an analytic truth.

Rubicon. Leibniz's commitment to the idea that free actions must be determined falls out of his commitment to the principle of sufficient reason. And Wolff endorsed this idea and argued that it can be derived from the principle of contradiction.[56] According to this line of reasoning, free actions are logically or "absolutely" contingent but hypothetically necessary.

Dominant as the Wolffian view may have been in Germany, Kant was not a fan of this distinction.[57] In his earliest metaphysical work, *New Elucidation of the First Principles of Metaphysical Cognition* (1755), Kant sides with Crusius ("their illustrious opponent"), arguing that if an action is hypothetically necessary, then its logical contingency does nothing meaningful to preserve the agent's ability to do otherwise:

> What is at issue is the necessitating principle: namely, *whence* the thing is necessary. I readily admit that here some of the adherents of the Wolffian philosophy deviate somewhat from the truth of the matter. They are convinced that that which is posited by the chain of grounds which hypothetically determine each other still falls a little short of complete necessity, because it lacks absolute necessity. But in this matter I agree with their illustrious opponent: the distinction, which everyone recites parrot-fashion, does little to diminish the force of the necessity or the certainty of the determination. For just as nothing can be conceived which is *more true* than *true,* and nothing *more certain* than *certain,* so nothing can be conceived which is *more determined* than *determined*. (*New Elucidation* 1:400)

Kant thinks that if an action is determined by hypothetical necessity, then the contingency that consists in the logical possibility of not performing the action means very little for the agent. But at this early point in his thought, he does not think anything important rides on our ability to do otherwise. Indeed, he wholeheartedly endorses Leibniz and Wolff's notion of spontaneity, and he does not seem to have any problem with the lack of contingency:

---

[56] For discussion of this argument, see Beck, *Early German Philosophy*, 256ff.

[57] By framing the *New Elucidation* as anti-Leibnizian on this count, my view is closer to that of Jeremy Byrd's, "Kant's Compatibilism in the *New Elucidation of the First Principles of Metaphysical Cognition*" and it parts ways with Beatrice Longuenesse, "Kant's Deconstruction of the Principle of Sufficient Reason." Although there are many affinities between Kant's view and Leibniz's, in the *New Elucidation*, Kant does not seem willing to admit that agents could have done otherwise or that this was a necessary condition of freedom.

> If, as happens in the case of machines, intelligent beings were to comport themselves passively in relation to those things which impel towards certain determinations and changes, I would not deny that the blame for all things could be shifted to God as the Architect of the machine. But those things which happen through the will of beings endowed with understanding and the spontaneous power itself of self-determination obviously issue from an inner principle, from conscious desires and from a choice of one of the alternatives according to the freedom of the power of choice. (*New Elucidation* 1:404)

The fact that the choice is not externally coerced but is instead determined by an "inner principle" is enough to secure its spontaneity. Of course, this is the very concept of spontaneity that he later rejects.

By the 1770s, Kant had fully abandoned that idea the idea that this kind of spontaneity was sufficient for meaningful human freedom. It is in this decade that the aforementioned "automaton" and "turnspit" criticisms begin to make an appearance (*L_1* 28:267). At this time, he is also ready to claim unequivocally (though without a particularly strong argument) that human beings really do have transcendental freedom:

> The soul is a being which acts spontaneously, simply speaking <*simpliciter spontan*>; i.e., the human soul is free in the transcendental sense <*in sensu transcendentali*> . . . (*L_1* 28:267–78)

It is this shift in his thinking, more than any other, that marks the most significant move toward the critical view of freedom. He had begun to think that the soul is *absolutely* spontaneous and does not exhibit a merely relative spontaneity. This move was largely indebted to Crusius, who freed Kant from the constraints of the Wolffian view. Kant no longer thought that every choice must be completely determined in accordance with the principle of sufficient reason: the free choices of intelligible beings are exempt. This was made possible by Crusius who argued that the Wolffian philosophy makes the mistake of conflating conditions of cognition for conditions of things. With this distinction in hand, Kant was ready to make the critical departure. If the relation of cause to effect holds strictly in virtue of the understanding, which imposes this concept on

objects of representation, then there is room for absolute spontaneity in the realm of things in themselves. The possibility of a spontaneous, uncaused cause is a live one as long as this cause is thought outside the realm of the world of appearances. In the next chapter I turn my attention to the first *Critique* in order to better understand Kant's argument for this conclusion.

Chapter 2:
Freedom in the Theoretical Philosophy

2.1 Introduction

Although freedom eventually becomes a central issue of Kant's practical philosophy, it makes its first appearance in the critical system as a cosmological problem within his theoretical philosophy. It arises out of a conflict between the understanding's firm commitment to determinism and the fact that reason is inevitably drawn to the belief that a complete explanation of the world requires the existence of a free, uncaused cause. Reason finds itself pulled in opposite directions. On the one hand, everything that happens in the phenomenal world takes place "solely in accordance with the law of connection of cause and effect" (B232). The concept of causality (which has its seat in the understanding) is constitutive of our representations of events that take place in the spatiotemporal world, and we know a priori that every event is causally necessitated by the laws of nature and the preceding state of affairs. Reason, on the other hand, does not limit itself to the domain of possible experience; it is compelled to extend "beyond the boundaries of experience" and consider the world as a whole (A296 /B353). And this is what leads us into metaphysical speculation about God and free will.

Reason overextends in this way because it is driven by a demand for a complete, systematic explanation of the world. This demand ultimately manifests itself in the form of the following principle: "If the conditioned is given, then the whole sum of conditions, and hence the absolutely unconditioned, is also given, through which alone the conditioned was possible" (A409 / B436). This is the principle that leads reason to posit the existence of freedom by means of the following argument. Whenever we encounter an event in the phenomenal world we know that it was conditioned by a previous event. But that event in turn must have also been conditioned by a

previous event. This regress continues ad infinitum. Like Aristotle, however, Kant argues that reason cannot abide this infinite regress because it cannot achieve "completeness of the series of the side of causes" (A446 / B474). Thus he concludes that "a causality must be assumed through which something happens without its cause being further determined by another previous cause, i.e., an absolute causal spontaneity beginning from itself a series of appearances that runs according to natural laws, hence transcendental freedom, without which even in the course of nature the series of appearances is never complete on the side of the causes" (A446 / B474).

Of course, this suggestion does not resolve the conflict; on the contrary, it is the root of the problem. Although positing transcendental freedom satisfies reason's demand for completeness in the series of causes, it also threatens to undermine the unity of experience, eliminating our knowledge of nature, as it would "cut us loose from the laws of nature and deprive us of that guiding thread of rules which permits us to establish the order and connection of events in the world."[58] The wellspring of all scientific knowledge is the principle that every alteration takes place in accordance with the law of cause and effect, and transcendental freedom threatens to undermine that principle.

Kant argues that this conflict is both natural and inevitable, but he also thinks that reason must find a resolution.[59] Our knowledge of the natural world and the unity of reason hang in the balance. But the situation is not hopeless; Kant claims that transcendental idealism should be able to rescue reason from this predicament. By distinguishing between appearances and things in themselves, Kant believes that he can provide a satisfactory resolution to the antinomy. He argues

---

[58] Carnois, *The Coherence of Kant's Doctrine of Freedom*, 3.

[59] Even after the root of the error has been exposed, Kant thinks that the conflict is never *fully* eradicated. The dialectic of reason is in this sense inevitable. "There exists, then, a natural and unavoidable dialectic of pure reason — not one in which a bungler might entangle himself through lack of knowledge, or one which some sophist has artificially invented to confuse thinking people, but one inseparable from human reason, and which, even after its deceptiveness [*Blendwerk*] has been exposed, will not cease to play tricks with reason and continually entrap it into momentary aberrations ever and again calling for correction." (A298 / B355)

that the deterministic laws of nature (and the principle that every event has a cause) hold only for objects of appearance, for they alone are subject to the schematized concept of causality, which the understanding imposes on its objects of representation. Things in themselves, by contrast, are not subject to the laws of nature and thus Kant thinks that the noumenal self can indeed possess absolute causal spontaneity. Most ambitiously, he thinks he can show that the free, noumenal causes could possibly have effects in the phenomenal world in a way that does not undermine the determinism of nature.

In this chapter, I show why Kant thinks this conflict between determinism and freedom inevitably arises and how he thinks transcendental idealism can resolve it. I begin in section 2.2 with a general overview of the Transcendental Analytic. That is where Kant argues for the claim that concepts like causality attach to all and only objects in the spatiotemporal world. I turn to the Dialectic in section 2.3 where I explain the argument behind reason's "demand for the unconditioned" and I demonstrate how this principle gives rise to the Third Antinomy. I conclude in section 2.4 where I argue that transcendental idealism mitigates the conflict but that the resolution has a limited scope. Theoretical reason cannot justify the assertion that humans possess transcendental freedom; it cannot even prove that freedom is metaphysically possible. Kant claims to have shown only that freedom is not logically impossible and that it is compatible with a commitment to determinism. The view I defend in this chapter is importantly distinct from many of those in the literature in that I do not commit Kant to any particular stance on how transcendental freedom works. On my reading, Kant does not have any positive view whatsoever of how freedom works; he offers only an explanation of how freedom and determinism could *potentially* be consistent with one another, not how they *actually* are. This chapter is thus central to my overarching aim concerning the incomprehensibility of freedom, for if Kant were committed to

something stronger than the view I articulate here, then he might not be able to defend the claim that we cannot comprehend how freedom is possible. In chapter 4, I explain what it is Kant thinks we cannot comprehend about free will, and that conclusion will depend on the view I defend here.

2.2 The Terra Firma of the Transcendental Analytic

In order to fully understand reason's demand for the unconditioned, it should be viewed in the broader context of Transcendental Dialectic. The Dialectic in turn must be seen in relation to the Transcendental Analytic. Thus I begin this section with a brief overview of the structure of the Analytic paying special attention to the concept of causality, both how it applies *and is restricted to* objects of appearance.

Kant begins with the Transcendental Aesthetic where he argues that space and time are neither real entities, nor mere relations;[60] rather, they are forms of sensibility which belong only to the "subjective constitution of our mind" (A23 / B37). This conclusion from the Aesthetic provides the foundation upon which the rest of the critical apparatus is to be built. It is here that Kant introduces both the distinction between things in themselves and appearances *and* the insight that the mind plays a role in giving form to the objects of appearance.[61]

With the Aesthetic behind him, Kant turns to what he calls the 'Transcendental Logic' in which he intends to describe the remaining contributions that the mind makes with respect to our cognition of objects. The distinction between the Transcendental Aesthetic and the Transcendental Logic arises out of one of the many dualisms of Kantian epistemology. As I explained in the

[60] Kant saw these two positions, belonging to Newton and Leibniz respectively, as the only rivals to his own view. The argument can be read as a disjunctive syllogism. Kant thinks that these are the only three options and that the Leibnizian and Newtonian positions fail. This has led to a famous criticism that there may be a neglected alternative, namely space and time might be the sensible form of our intuition and *also* be properties of things in themselves.

[61] The distinction is discussed earlier in the preface but it is more fully explained and defended in the Aesthetic.

previous chapter, the **sensibility** is receptive and representations are given to it. The **understanding** is our faculty for thinking about the objects received through the sensibility. Cognition depends on the cooperation of these faculties. This is the idea behind one of the most famous slogans of the first *Critique*: "Without sensibility no object would be given to us, and without understanding none would be thought. Thoughts without content are empty, intuitions without concepts are blind" (A51 / B75).

The task of the Transcendental Logic is to outline the workings of the understanding and to explain how its concepts and principles apply to the objects of cognition. Kant begins with an analysis of the pure forms of thought, and he delineates the forms that judgments can take "[i]f we abstract from all content of a judgment in general, and attend only to the mere form of the understanding in it" (A70 / B95). Kant's ideas about logic are largely Aristotelian. The idea here is that we can remove all content from a judgment and examine its pure, logical form.

After providing an exposition of the pure forms of judgment (what Kant calls 'general logic,') he moves to the question of the form that these judgments must take when applied to intuitions that are given to it through the sensibility. Each form of judgment within the general logic has a counterpart in the table of categories in the transcendental logic. For instance, the logical relation expressed by the conditional manifests itself as the relation of cause and effect when it is applied to objects of cognition. Proceeding in this fashion, (converting each form of judgment to a category of the understanding), Kant arrives at the table of categories, which he believes to be an exhaustive list of the a priori concepts which are required for cognition.

The presentation of the categories of the understanding is followed by a section which contains one of the most important arguments of the *Critique* — the Transcendental Deduction. Unfortunately, the argument of the Deduction is notoriously as difficult as it is important. Indeed,

even Kant underscores both of the obscurity and the centrality of the Deduction.[62] His aim in the Deduction is to demonstrate how the a priori categories of the understanding can legitimately apply to objects of experience. Like many of the arguments in the *Critique*, the Deduction can be understood as a response to the skeptical challenges put forth by Humean empiricism.[63] Hume argued that our experiences cannot justify the assertion that there is a necessary connection between causes and effects, and he thinks that we should therefore reject the objective validity of the concept of causality. As Hume sees it, we might continue to employ the idea as a matter of custom or habit, but the idea of necessary connection does not properly belong to the objects of experience.

Kant concedes to Hume that experience alone (in Hume's sense) cannot justify a belief in causal necessity, but he rejects the conclusion that this nullifies the necessary connection between causes and effects. On the contrary, Kant thinks that the concept of causality not only applies to objects of experience but that it *necessarily* applies to them. The full details of the argument need not trouble us here as they lie outside the scope of my project.[64] But the conclusion is central to

---

[62] He writes, "I am acquainted with no investigations more important for getting to the bottom of that faculty we call the understanding, and at the same time for the determination of the rules and boundaries of its use, than those I have undertaken in the second chapter of the Transcendental Analytic, under the title **Deduction of the Pure Concepts of the Understanding**; they are also the investigations that have cost me the most, but I hope not unrewarded, effort" (A xvi). Kant was disappointed to discover that his costly labors did indeed go unrewarded as readers were universally baffled by the Deduction. Because of the negative response, Kant was compelled to completely rewrite the Deduction in the second edition. This was one of the only sections to undergo this kind of revision. Kant laments the fact that the deduction involves such "inevitable obscurity" (A98).

[63] In his discussion of the Deduction, Derk Pereboom frames the argument in this way. See Pereboom, "Kant's Metaphysical and Transcendental Deductions."

[64] Here is a brief summary of the Deduction. In its most simplified form, it boils down to two premises. The first premise concerns self-consciousness, more specifically the transcendental unity of apperception. Kant claims that "It must be possible for the "I think" to accompany all my representations" (B 131). Contrary to Hume's conclusion that we are only aware of a bundle of experiences, Kant is arguing that we necessarily posit a source of this synthetic activity: "This thoroughgoing identity of the apperception of a manifold which is given in intuition contains a synthesis of representation, and is possible only through the consciousness of this synthesis" (B 133). The fact that we must attach the "I think" to all of our representations is evidence that the subject plays a vital role in synthesizing experience.

The second premise is that the synthetic activity which is at stake is objective and not merely subjective. (B139). Although empirical self-consciousness is contingent and subjective, the transcendental unity of apperception is necessary and objective. Pereboom points out how this argument foreshadows the Second Analogy ("Kant's

my aim, since the concept of causality is one of the categories at stake in the Deduction. The upshot is two-fold. First, "We cannot **think** any object except through the categories; we cannot **cognize** any object that is thought except through intuitions that correspond to these concepts" (B165). That is to say, *all* objects of experience are subject to the categories and we cannot think of them in any other way. Second, our cognition "is limited to objects of experience" (B166). The concepts can yield no cognition without an object: "for if an intuition corresponding to the concept could not be given at all, then it would be a thought as far as its form is concerned, but without any object, and by its means no cognition of anything at all would be possible . . . all intuition that is possible for us is sensible (Aesthetic), thus for us all thinking of an object in general through a pure concept of the understanding can become cognition only insofar as this concept is related to an object of the senses" (B147). This is the second half of the conclusion. It was just seen above that *all* objects of experience must be cognized through the categories, but now the claim is that *only* objects of experience can be cognized through the categories.[65]

The understanding can synthesize only what is given to it by the sensibility. And sensibility is constrained by the fact that we receive intuitions through the forms of space and time. Therefore, we can cognize only spatiotemporal objects. Things in themselves are entirely outside the bounds of possible experience. When it comes to Kant's theory of transcendental freedom, the importance of this conclusion, which comes up several times in this chapter, cannot be overstated. This is what

---

Metaphysical and Transcendental Deductions,"163). The synthesis performed by the understanding is done by means of the categories (as it was argued earlier that these categories are complete and exhaustive). Thus, Kant concludes the categories have objective validity. As Strawson puts it, "the fact that my experience is of a unified objective world is a necessary consequence of the fact that only under this condition could I be conscious of my diverse experiences as one and all my own" (*Bounds of Sense*, 94).

[65] It is important to note (as Kant does in a footnote) that this does *not* mean that the categories are meaningless outside the context of cognition. As I explain below, we can legitimately "think" about things in themselves through the unschematized categories. Without this addendum Kant's claim that things in themselves ground appearances would be inconsistent with the claim that that the category of causality and dependence has no application outside the world of experience.

will allow Kant to argue that although causality necessarily pertains to the spatiotemporal world, we do not have any reason to think that things in themselves are constrained in the same way.

Kant elaborates on this conclusion in a short but important section that follows the Deduction — the schematism chapter. Kant points out that there is a missing link between the concepts of the understanding and objects of appearance. He believes that a third thing is needed to mediate between the pure categories and appearances. For a contrasting case (i.e., one where no third thing is needed), he uses the example of the concept of a dinner plate and the concept of a circle. He says that the "roundness that is thought in the former can be intuited in the latter" (A137 / B176). The two concepts are thus "homogenous." The problem for Kant's claim about pure a priori concepts attaching to objects of experience is that "the pure concepts of the understanding" (e.g. the concept of causality) and empirical intuitions "are entirely un-homogenous" (A137 / B136). As Kant puts it, "there must be a third thing, which must stand in homogeneity with the category on the one hand and the appearance on the other, and makes possible the application of the former to the latter" (A138 / B177).

He adds a further constraint on what this third thing must be, as he says that the "mediating representation must be pure (without anything empirical) and yet **intellectual** on the one hand and **sensible** on the other." The solution to the puzzle turns out turns out to be time. Time is homogenous with the "**category** (which constitutes its unity) insofar as it is universal and rests on a rule a priori. But it is on the other hand homogenous with the **appearance** insofar as **time** is contained in every empirical representation of the manifold. Hence an application of the category to appearances becomes possible by means of the transcendental time-determination which, as the schema of the concept of the understanding, mediates the subsumption of the latter under the former" (A138 / B177).

The pure concepts of the understanding (causality being the most important for my purposes here) can become meaningful for us only if an object is given. And, given the forms of sensibility, we know that this requires temporality. Without an object, the concepts are empty. Kant says that "concepts are entirely impossible,[66] and cannot have any significance, where an object is not given for themselves or at least for the elements of which they consist, *consequently they cannot pertain to things in themselves*" (A139 / B 178; emphasis added). He refers to this condition as the "schema" of the understanding and this procedure through which the categories acquire meaning is called the "schematism." It is only within the domain of possible experience that concepts like causality can have any concrete meaning or sense (*Sinn*).[67]

Because of this limitation, the schematized concept of causality is restricted to the world of phenomena. Nevertheless, Kant says that we are naturally drawn to think about things in themselves through the categories even though no object is given:

> Now if we leave aside a restricting condition, it may seem as if we amplify the previously limited concept; thus the categories in their pure significance, without any conditions of sensibility, should hold for things in general, **as they are**, instead of their schemata merely representing them **how they appear**, and they would therefore have a significance independent of all schemata and extending far beyond them. In fact, even after abstraction from all sensible condition, significance, but only a logical significance of the mere unity of representations, is left to the pure concepts of the understanding, but no object and thus no significance is given to them that could yield a concept of the object. Thus, e.g., if one leaves out the sensible determination of persistence, substance would signify nothing more than a something that can be thought as a subject (without being a predicate of something else). Now out of this representation I can make nothing, as it shows me nothing at all about what determinations the thing that is to count as such a first subject is to have. Without schemata, therefore, the categories are only functions of the understanding for concepts, but do not represent any object. This significance comes to them from sensibility, which realizes the understanding at the same time as it restricts it. (A146–47 / B186–87)

---

[66] In his copy of the first edition, Kant changed 'impossible' to 'are for us without sense.' See page 272 of the Cambridge edition of the *Critique*.

[67] Eric Watkins points out that the possibility of experience "consists of three elements: inner sense (along with its a priori form, time), the imagination's synthesis of representations (in inner sense), and the unity of apperception's synthetic unity (in concepts and judgment)" ("System of Principles," 152).

We are tempted to think that the categories can give us knowledge of things in themselves in the same way that they allow us to cognize objects of experience. But the understanding cannot accomplish this because its concepts have only a bare logical significance in the absence of a given object.

The chapter which follows the schematism contains one of Kant's most important discussions of causal necessity: the Second Analogy. This is where Kant puts forth the principle that "All alterations occur in accordance with the law of cause and effect" (B232).[68] Kant once again emphasizes the point that this causal principle and the other principles of the understanding "are nothing other than principles of the determination of the existence of appearances in time" (A215 / B262). Through this principle, we understand that nature is a unified whole in that everything which happens in the spatiotemporal world of appearances is determined by laws. But, crucially, the principle of causal necessity is limited to the world of phenomena: "we cognize only the necessity of **effects** in nature, the causes of which are given to us, and the mark of necessity in existence does not reach beyond the field of possible experience" (A227 / B280).

The final chapter of the Transcendental Analytic is titled "*On the ground of the distinction of all objects in general into* phenomena *and* noumena" (A235 / B294). Yet again, Kant repeats the claim that the categories are to be applied only within the empirical world. But this limitation means that the understanding is not able to ponder the ultimate sources of its cognition: the noumenal world. This is one of the places where Kant waxes poetic:

> We have now not only traveled through the land of pure understanding, and carefully inspected each part of it, but we have also surveyed it, and determined the place for each thing in it. But this land is an island, and enclosed in unalterable boundaries by nature itself. It is the land of truth (a charming name), surrounded by a broad and stormy ocean, the true seat of illusion, where many a fog bank and

[68] This is the version he gives in the second edition. In the first edition the principle is worded differently. It reads, "Everything that happens (begins to be) presupposes something which it follows in accordance with a rule" (A 189).

> rapidly melting iceberg pretend to be new lands and, ceaselessly deceiving with empty hopes the voyager looking around for new discoveries, entwine him in adventures from which he can never escape and yet also never bring to an end. (A235 / B294).

After his thorough survey of the understanding, Kant is ready to embark on a voyage onto the broad and stormy ocean of illusion. But he takes one last look at the terra firma of the understanding to see if the *Critique of Pure Reason* is truly compelled to journey into the sea of illusion:

> Before we venture out on this sea, to search through all its breadth and become certain of whether there is anything to hope for in it, it will be useful first to cast yet another glance at the map of the land that we now leave, and to ask, first, whether we could not be satisfied with what it contains, or even must be satisfied with it out of necessity, if there is no other ground on which we could build; and, second, by what title we occupy even this land, and can hold it securely against all hostile claims. (A236 / B295).

The understanding can inform us well enough about the objects of experience. All of our scientific knowledge springs from this source. The entire edifice is built on the firm foundation of experience. But this strength is also something of a weakness. Since the understanding can consider only the world of experience, it can know nothing about its noumenal source. And if the *Critique* is to complete its task, then Kant must investigate the noumenal world in whatever capacity reason permits him.

The first question is how, if at all, we come to know about the noumenal realm. Once more, Kant makes one of his favorite distinctions between a positive and negative sense: "If by a noumenon we understand a **thing insofar as it is not an object of our sensible intuition**, because we abstract from the manner of our intuition of it, then this is a noumenon in the **negative** sense. But if we understand by that an **object of a non-sensible intuition**, then we assume a special kind of intuition, namely intellectual intuition, which, however, is not our own, and the possibility of which we cannot understand, and this would be noumenon in a **positive** sense" (B307). Kant says

very little about intellectual intuition, but it is important to note that whatever it is like,[69] it is entirely closed off to us. We can cognize only what is given to us through sensible intuition, and the form of that intuition is necessarily spatiotemporal. So although God might be able to have intellectual (i.e., non-sensible) intuition of things in themselves (and thus cognize noumena in the positive sense), we cannot.

The negative sense, however, is very important for the *Critique*. Indeed, this concept is indispensable. As early as the preface, Kant points out that although we cannot "cognize" things in themselves, we must "think" about them. The world of appearance requires a source: "Yet the reservation must also be well noted, that even if we cannot cognize these same objects as things in themselves, we at least must be able to think them as things in themselves. For otherwise there would follow the absurd proposition that there is an appearance without anything that appears" (B xxvi). We are required to think that there must be a thing in itself even though we cannot cognize it or have any intuition of it. Kant says that "it is therefore merely a **boundary concept**" that we put forward in order to make sense of the claim that there must be something behind our appearances (A255 / B 311). He says that "a concept of a *noumenon*, taken merely problematically,[70] remains not only admissible, but even unavoidable, as a concept setting limits to our sensibility" (A256 / B311). Although Kant is about to journey into the land of illusion in

[69] The concept of intellectual intuition is fairly thin in Kant. It would come into greater currency with the later German Idealists, especially Fichte who argues that intellectual intuition is required for the 'I.' But Kant makes very few references to it, except to say that we cannot have this kind of intuition. We can only receive intuitions through the forms of space and time. God, by contrast, would be able to intuit objects immediately, without receiving them through the sensibility. As Julian Wuerth describes it, intellectual intuitions are *creative* in that "thoughts match their objects by virtue of creating them" ("Paralogisms of Pure Reason," 213).

[70] Kant is using 'problematic' here as a technical term. It is one of his categories of modality. Kant makes several references to his idea of "real possibility" in this chapter. The idea is that the non-contradictory nature of a concept is not enough to secure its real (i.e., metaphysical) possibility. If a concept contains no contradiction, then Kant thinks it is logically possible, but it might still be metaphysically impossible. Cognition is the only way that real possibility can be secured. Since we have no cognition of noumena, we cannot know that they are *really* possible. They are logically possible, but we must remain agnostic about their real possibility (at least until the second *Critique* where practical cognition allows us to have some insight into the noumenal realm). I will use 'problematic' in Kant's sense throughout the dissertation and I usually add quotation marks to indicate the technical usage.

the Transcendental Dialectic, he uses this final chapter of the Analytic to reinforce the limits of the understanding. Noumena cannot be cognized through the categories, reason is "merely thinking them under the name of an unknown something" (A256 / B311).

2.3 Setting Sail on the Sea of Illusion

Kant begins the Transcendental Dialectic with a somewhat puzzling discussion of "transcendental illusion." Commentators have long been troubled by an apparent inconsistency that runs through the entire Dialectic. On the one hand, Kant seems to think that the critical machinery of transcendental idealism should put an end to the contradictory cosmological claims of the antinomies. In this way, the *Critique* is supposed to censure our predilection for wanton metaphysical speculation and lay it to rest. But, on the other hand, he repeatedly emphasizes that the transcendental illusion is "inevitable" and that it "does not cease even though it is uncovered and its nullity is clearly seen into by transcendental criticism" (A297 / B353). The *Critique* "can never bring it about that transcendental illusion (like logical illusion*) should ever disappear and cease to be an illusion.*"[71] How can both of these claims hold? If transcendental idealism resolves the contradictions, then why doesn't the illusion cease? Conversely, if the contradictions are inevitable and unceasing, then what good does it do to invoke transcendental idealism?[72] In this chapter I propose a reading of the Dialectic that offers a way out of this dilemma. First, I explain how transcendental illusion inevitably arises and how the conflict between freedom and

[71] *Ibid.* Emphasis added.

[72] My discussion of transcendental illusion is indebted to Michelle Grier's excellent book on the topic: *Kant's Doctrine of Transcendental Illusion*. She points out that this accusation of inconsistency is rampant in the literature. Patricia Kitcher also levels this criticism in "Kant's Paralogisms," 518. But it appears in many places. Cf. Walsh, *Kant's Criticisms of Metaphysics*; Norman Kemp Smith, *A Commentary to Kant's* Critique of Pure Reason, 457; Strawson, *The Bounds of Sense,* 155–61.

determinism emerges in this context. In the following section, I explain how (and to what extent) transcendental idealism can mitigate our metaphysical maladies.

Kant begins by contrasting transcendental illusion with logical illusion. The main difference between the two is that putting an end to logical illusion is straightforward. Logical illusion results from a simple "failure of attentiveness to the logical rule" (A296 / B353). A student might be fooled into thinking that P follows from the premises 'If P then Q' and 'Q.' But once she comes to understand the difference between modus ponens and affirming the consequent, she will see that P does not follow from those premises. Transcendental illusion is harder to eradicate because it does not rest simply on a logical fallacy.[73] It "irremediably attaches to human reason, so that even after we have exposed the mirage it will not cease to lead our reason on with false hopes, continually propelling it into momentary aberrations that always need to be removed" (A298 / B355). Since transcendental illusion has its seat in pure reason, I begin this section with an explanation of what Kant means by pure reason.

2.3.1 Pure Reason as the Seat of Transcendental Illusion

As usual, Kant begins with a distinction. He distinguishes between the logical and pure uses of reason. The logical use is "where reason abstracts from all content of cognition" (A299 / B355). As discussed earlier, this is Kant's description of formal logic, and it is the syllogistic reasoning that he inherits from Aristotle: "In every syllogism I think first a **rule** (the *major*) through the understanding. Second, I **subsume** a cognition under the condition of the rule (the *minor*) by means of the **power of judgment**. Finally, I determine my cognition through the predicate of the

[73] Nevertheless, the dialectical inferences *do* contain fallacies. It's just that the fallacies that arise out of transcendental illusion result from a natural and inevitable conflation of objective and subjective conditions. So although logical illusion ceases when the fallacy has been exposed, transcendental illusion remains as we are prone to fall back into the same error.

rule (the *conclusio*), hence *a priori* through **reason**" (A304 / B360). Kant notes, however, that this process often works in the other direction. "[A]s happens for the most part, the conclusion is a judgment given as the problem" and the task then is to find a universal rule and ask whether or not the object in question can be subsumed under that rule. For instance, we might start from a judgment like "this cup is heavy," and then we work our way back to a universal rule, subsuming the cognition under it: "All bodies are heavy."[74] "This cup is a body." By using this process, reason's knowledge becomes increasingly general. Rather than just having a piece of knowledge about this particular object, we have learned a universal condition that extends to other objects. Kant thinks that this is one of the principal virtues of reason. It amplifies our knowledge by making it more general, systematic, and unified. He says, "Now if I find such a condition and if the object of the conclusion can be subsumed under the given condition, then this conclusion is derived from the rule that **is also valid for other objects of cognition.** From this we see that reason, in inferring, seeks to bring the greatest manifold of cognition of the understanding to the smallest number of principles (universal conditions), and thereby to effect the highest unity of that manifold" (A305 / B361). This is what Kant calls the "logical use of reason." We want to find the smallest set of principles for the largest possible unity.

And in addition to the logical use, "there is also a real use" which comprises reason's capacity to put forth concepts and principles "which it derives neither from the senses nor the understanding" (A299 / B355). These principles of reason must be sharply distinguished from the principles of the understanding. For instance, as we saw above, the understanding is the seat of the principle that "everything that happens has a cause," and such a principle confers unity on appearances. The principles of reason by contrast do not take any experience as their object; rather,

[74] In a different context, Kant does assert this universal condition for bodies: "All bodies are heavy" (A7 / B11).

the principles of reason apply to the understanding: "If the understanding may be the faculty of unity of appearances by means of rules, then reason is the faculty of the unity of the rules of understanding under principles. Thus it never applies directly to experience or any object, but instead applies to the understanding, in order to give unity *a priori* through concepts to the understanding's manifold cognitions, which may be called 'the unity of reason' (A302 / B359). This point of contrast is important. It explains the way in which reason and the understanding have separate domains. The principles of the understanding (e.g., the principle that every event has a cause) apply to objects. They constitute the nature of those objects. It is only by means of those concepts that we are able to cognize objects. The metaphysical principles of "pure reason" do *not* apply to objects in this way.

Nevertheless, like the understanding, reason is concerned with unity. Reason is consumed with the pursuit of a certain kind of parsimony. It wants to find the broadest and most universal conditions.[75] In order to unify the understanding, reason seeks increasingly general conditions. Whenever we discover a universal rule (the major premise) which allows us to subsume a cognition (the minor premise), we notice that "this rule is once again exposed to this same attempt of reason, and the condition of its condition thereby has to be sought" (A307 / B364). This means that we do not stop at the discovery that "this event has a cause." For every condition that we discover, we seek a more general condition. We want a condition for *the entire series of conditions*.[76] Reason

[75] Kant makes direct reference to the "economy" of principles as he says that reason "does not prescribe any law to objects, and does not contain the ground of the possibility of cognizing and determining them as such in general, but rather is merely a subjective law of economy for the provision of our understanding, so that through comparison of its concepts it may bring their universal use to the smallest number" (A306 / B362–63).

[76] It might be helpful to understand this principle by comparing it to Spinoza's idea of God as the immanent cause of all things. Spinoza claims that "God is the immanent, not the transitive, cause of all things" (*E* IP18). There exists a transitive series of causes: A causes B, which causes C, which causes D, etc. For Spinoza, God is not a cause in this sense. Rather, God is the cause of the entire series of causes in that all things are in God and follow from God's nature. For Kant, the situation is similar. For any given event, we know that it had a cause preceding it in time. But the entire series of conditions *also requires a condition*. On Spinoza's view, this is God. For Kant, it is the "unconditioned

is after the most general universal conditions, and thus its search will end only when it arrives at "the unconditioned for conditioned cognitions" (A307 / B364).

The change in wording is crucial. The "proper principle of reason (in its logical use) is to **find** the unconditioned for conditioned cognitions of the understanding, with which its unity will be completed" (A307 / B364; emphasis added). The "unconditioned" is given to us as a **task**; we are supposed to *seek* the unconditioned. But this legitimate principle of "logical reason" takes a special (and more pernicious) form in the context of "pure reason." Kant says that "this logical maxim cannot become of a principle of **pure reason** unless we assume that when the conditioned is given, then so is the whole series of conditions subordinated one to the other, which is itself unconditioned, also given (i.e., contained in the object and its connection)" (A307 / B364). Reason is moving from one principle to another:

> Principle of Logical Reason (**PLR**): Seek the unconditioned for conditioned cognitions.[77]
>
> Principle of Pure Reason (**PPR**): If the conditioned is given, then the whole sum of conditions, and hence the absolutely unconditioned, is also given.[78]

This is a surprising result. Far from the logical practice of **seeking** universal conditions for particular cognitions, Kant thinks reason has now led us to adopt a metaphysically extravagant principle which tells us that whenever a conditioned thing is given, the unconditioned is **given** as well. This is what gets reason into trouble. It is because of our assumption that the unconditioned is *given* that we come to speculate about things which transcend the boundaries of experience. We become tempted to think that things like God and the soul are given, even though they can never

---

condition." The particular referent of the 'unconditioned condition' depends on the context. In the Third Antinomy it is a free cause. In the Fourth Antinomy it is a necessary being.

[77] Kant's wording for the principle: "Find the unconditioned for conditioned cognitions of the understanding, with which its unity will be completed" (A308 / B364).

[78] A409 / B436

be encountered in experience. Thus, Kant is quick to warn us that the ideas which arise from this "supreme principle of pure reason" are **transcendent** with respect to appearances; whereas the principles of the understanding are **immanent.**

The principles "whose application stays wholly and completely within the limits of possible experience" are **immanent**, and those that "would fly beyond these boundaries" are **transcendent** (A295 / B352). The principle that "every event has a cause" is immanent as it applies only to objects of experience. In this sense, immanent principles are law abiding; they remain safely within the boundaries of experience. Transcendent principles, however, "incite us to tear down all those boundary posts and to lay claim to a wholly new territory that recognizes no demarcations anywhere" (A296 / B352). Hence, Kant's admonition is that the **PPR** has no bearing on the world of experience: "no adequate empirical use can ever be made of that principle." But with this warning in mind, we might wonder what utility this principle has, since it has no empirical value and it is this principle that inevitably leads reason into a web of contradictions.

The answer is that Kant thinks without such a principle our cognition would lack unity. He says that "All our cognition starts from the senses, goes from there to the understanding, and ends with reason, beyond which there is nothing higher to be found in us to work on the matter of intuition and bring it under the highest unity of thinking" and he calls pure reason the "supreme faculty of cognition" (A298 / B355). Consider, for instance, the idea of the world as a unified totality. The understanding is not able to cognize anything other than what is given to it through the sensibility. Thus it arrives at one of its boundaries. Experience does not give us any cognition of "the world" as a whole; we never have an intuition of the entire totality of objects (A483 / B511). Experience of objects comes to us in a piecemeal fashion. But Kant thinks this is a problem for the completeness of our knowledge. When we are working on discovering the laws of physics,

for example, we must think that these laws apply to the entire world, not just to the handful of phenomena that we have observed. We assume that our experiences are systematically connected to a totality (even though we have no experience of this totality as such).[79] Although reason's supreme principle (**PPR**) can indeed get us into some trouble when it is misapplied, the completeness of our knowledge requires it. Reason "necessarily and with every right demands" the unconditioned" (Bxx).

The problem is that reason's demand for the unconditioned does not stay within its boundaries. Properly speaking, it is "only a logical prescription in the ascent to ever higher conditions to approach completeness in them and thus to bring the highest possible unity of reason into our cognition," but "this need of reason, has, through a misunderstanding, been taken for a transcendental principle of reason, which overhastily postulates such an unlimited completeness in the series of conditions in the objects themselves" (A309 / B365–66). Rather than restricting reason's demand for the unconditioned (to its proper domain as a subjective condition of thought), we mistakenly extend it as an objective condition in the realm of experience.

2.3.2 Transcendental Ideas and Dialectical Inferences

The most conspicuous consequence of the PPR is that it gives rise to certain ideas which do not correspond to any object of experience but which are nevertheless legitimate objects of

[79] Kant summarizes this idea nicely in his discussion of the regulative value of the cosmological idea of the world-whole: "For if the greatest possible empirical use of my reason is grounded on an idea (that of systematic complete unity, about which I will have more to say presently), which in itself can never be presented adequately in experience, even though it is unavoidably necessary for approximating to the highest possible degree of empirical unity, then I am not only warranted but even compelled to realize this idea, i.e., to posit for it an actual object, but only as a Something in general with which I am not acquainted at all and to which, as a ground of that systematic unity and in relation to that, I give such properties as are analogous to the concepts of the understanding in their empirical use" (A677 / B705).

thought. Kant calls them "transcendental ideas," and there are exactly three sources for them:[80] the thinking subject (the soul), the world (as the sum total of appearances), and God. There can only be three because they emerge from the three relations of judgment: categorical, hypothetical, and disjunctive. As he explains in the Analytic, there are three relations of thoughts: "(a) those of the predicate to the subject, (b) of the ground to the consequence, and (c) between the cognition that is to be divided and all the members of the division" (A73 / B98). When thinking about the unconditioned (as a "synthetic unity of all conditions") via these relations, reason generates its transcendental ideas. In the first case, the unconditioned qua subject /predicate manifests itself as the **soul**, "a subject that is no longer a predicate." The second idea — the **world** — is the unconditioned as the unity of all appearances. And the idea of **God** is the unconditioned as the "unity of the condition of all objects of thought in general" (A334 / B391).

Far from being mere phantoms of the mind, the transcendental ideas are necessary products of reason. Although the idea of the world as a totality can never be encountered in experience, it serves as a guiding light for our investigations in science. We learn about things in nature individually. We observe a particular thing at a particular time. But we must think of ourselves as being in the process of putting together pieces of a puzzle which could in theory become a complete picture of the world and its laws. We might observe an apple falling from a tree one day and the planets orbiting the sun the next day. But if scientific knowledge is to make any progress, then we must connect all of these disparate experiences and think of things in nature as parts of a unified world governed by universal laws. In this way, the transcendental ideas serve as a guiding light

[80] Kant sometimes refers to other concepts as transcendental ideas, which would mean that there might be more than three of them. Most important for my purposes is transcendental freedom, which he *does* call a 'transcendental idea.' He says "Freedom in this signification is a pure transcendental idea" (A533 / B561). But the transcendental idea of freedom arises out of our contemplation of the transcendental idea of the world. In that sense, it is a sub-species of the broader transcendental idea. So depending on the leniency of the definition, there might be more than three transcendental ideas. More precisely, there are exactly three *sources* of transcendental ideas.

that unifies our experiences. As Michelle Grier puts it, Kant's suggestion is that "these ideas are implicit in the practices governing scientific classification, and enjoin us to seek explanatory connections between disparate phenomena. As such, reason's postulations serve to provide an orienting point towards which our explanations strive, and in accordance with which our theories progressively achieve systematic interconnection and unity."[81]

Unfortunately, this positive aspect of reason's demand for the unconditioned is not the end of the story. There is a negative side which flows directly from the positive.[82] The negative manifestation of reason's demand for the unconditioned leads us to mistaken conclusions about the soul (paralogisms), entangles itself in contradictions about the world (the antinomies), and brings us into a field of unwarranted speculation about God (the ideal). For each transcendental idea, there is a species of dialectical error associated with it. In each case, we employ "syllogisms containing no empirical premises, by means of which we infer from something with which we are acquainted to something of which we have no concept, and yet to which we nevertheless, by an unavoidable illusion, give objective reality" (A339 / B397). It will become clear below that the transcendental ideas might be legitimate objects of thought, but we err disastrously when we regard them as *given* things with objective reality.

The first set of dialectical inferences concerns the thinking subject. Kant argues that the transcendental ideas are indispensable concepts of reason that are required for guiding our thoughts, and so it is with the idea of the soul. A crucial premise of the Transcendental Deduction was that we must be able to attach the "I think" to all of our representations; otherwise the representations would be nothing to us (B132). And so it is perfectly natural for us to want to know

[81] Michelle Grier, "Kant's Critique of Metaphysics."

[82] Unlike most of the other instances of the positive /negative distinction, this one is normatively loaded. The positive side (the transcendental ideas) are helpful for unifying our knowledge, but the negative side (the dialectical inferences) only lead us astray as we become prone to groundless metaphysical speculation.

more about this 'I' which is a necessary condition of our experiences. The other necessary conditions of experiences (space, time, and the categories of the understanding) were amenable to thorough investigation and those were so fruitful that they yielded the entire critical apparatus of the Analytic. Unfortunately, we cannot expect the same fecundity from our psychological investigations.

It should be noted first that Kant is not talking about empirical psychology. He thinks that empirical psychology is a perfectly legitimate enterprise. Empirical psychology takes as its subject matter the subject as it appears to itself (the phenomenal subject); it concerns neither the noumenal subject nor the transcendental 'I' of apperception. Nevertheless, Kant thinks there are very important questions that empirical psychology can answer. For instance, he spends a great deal of time wondering about the nature of human desire, pleasure, pain, etc. and such questions belong strictly to empirical psychology.[83]

Here in the Dialectic, however, Kant is concerned only with the "**rational** doctrine of the soul," and this must not contain "the least bit of anything empirical" (A342 / B400). The "**I think**" is thus the sole text of rational psychology" (A343 / B401). Kant thinks that several seemingly natural inferences follow from the "I think." Descartes famously begins with the *cogito* and derives a variety of metaphysical conclusions: that the soul is a thinking substance, that it is distinct from the material world, that it can survive the death of the body, etc. Kant agrees that these inferences seem perfectly reasonable at first glance, but he thinks that they all contain a crucial error.

Consider, for instance, the argument from the "I think" to the conclusion that the soul is a substance. This is the first paralogism that Kant discusses. The argument starts from the definition

[83] For a thorough discussion of Kant's thoughts on empirical psychology, see Patrick Frierson, *Kant's Empirical Psychology*.

of 'substance' as that which serves as the subject of judgments/predicates but which is not predicated or said of anything else. This is the classical, Aristotelian definition of substance.[84] The second premise is that the thinking subject (the soul) satisfies this definition. All thoughts and representations are predicated of the thinking subject, and it is not predicated of anything else. It is thus concluded that the soul is a substance and that various Cartesian properties follow: that the "I, as a thinking being, **endure** for myself, that naturally I **neither arise** nor **perish,**" that the soul is "everlasting through all alterations, even the human being's death" (A349–50). Kant thinks that this argument commits the fallacy of equivocation[85] as the meaning of 'substance' slides from the bare logical significance as the subject of predication to the more robust empirical category of substance. Recall that, for Kant, concepts of the understanding (such as substance) have no content if an object is not given. They take on concrete meaning only if a determinate object is given through intuition. The problem here is that no object is given. The 'I' that is contained in the "I think" is *not* an object of experience,[86] and so the category of substance cannot apply to it. Thus, the major premise employs the concept 'absolute subject' in the pure, logical sense, but the concept is then extended as if it had a determinate object.[87]

---

[84] This is the linguistic half of the definition. The ontological half is that a substance is that which everything else depends on for its existence but which does not depend on anything else. See Aristotle, *Categories*, Section 1, Part 5.

[85] Kant tends to use the Latin '*sophisma figurae dictionis.*' In keeping with Kant's Aristotelian logic, we might also call it the fallacy of ambiguous middle, since the argument relies on the ambiguity of the middle term.

[86] There are other concepts of the self in Kant, and some of these can be encountered in experience. The empirical self, (the self as it appears) is indeed an object of inner sense. The paralogisms, however, take the 'I think' as their only subject matter. But it's not as if the argument would fare any better when applied to the empirical self. Since the empirical self *is* an object of representation, it does *not* satisfy the first premise. It is predicated of something else. The self as it appears is not an 'absolute subject' in the sense that the major premise requires. Patricia Kitcher has written several illuminating commentaries on the paralogisms and Kant's concepts of the self. See Kitcher, "Kant on Self-Consciousness;" "Kant's Paralogisms;" *Kant's Thinker* and *Kant's Transcendental Psychology*. Cf. Marshall, "Kant's Metaphysics of the Self."

[87] Michelle Grier summarizes this point nicely: "the major premise deploys the term 'substance' in a very general way, one which abstracts from the conditions of our sensible intuition (space and time). As such, the major premise simply offers the most general definition of substance, and thus expresses the most general rule in accordance with which objects might be able to be thought as substances. Nevertheless, in order to apply the concept of substance in such a way as to determine an object, the category would have to be used empirically. Unfortunately, such an empirical

The next two paralogisms follow this structure as well.[88] We begin from the legitimate idea of the soul, but we are then led astray as we naturally want to treat this subject of pure thought as if it were a determinate object. In the conclusion, Kant describes the error as follows:

> The dialectical illusion in rational psychology rests on the confusion of an idea of reason (of a pure intelligence) with the concept, in every way indeterminate, of a thinking being in general. I think of my self, in behalf of a possible experience, by abstracting from all actual experience, and from this conclude that I could become conscious of my existence even outside experience and of its empirical conditions. Consequently I confuse the possible **abstraction** from my empirically determined existence with the supposed consciousness of a **separate** possible existence of my thinking Self, and believe that I cognize what is substantial in me as a transcendental subject, since I have in thought merely the unity of consciousness that grounds everything determinate as the mere form of cognition. (B426–27).

All that is contained in the 'I' of the 'I think' is a unity of consciousness, which, as we saw in the Deduction, is a necessary condition of experience. The danger is that we are wont to treat this idea as a determinate object and thus illegitimately attribute certain properties to it (endurance through time, personal identity, substantiality, etc.).

It would be reasonable to expect freedom to make an appearance in this section of the *Critique*. After all, freedom is a property of the self, and so it would seem that Kant's discussion of it belongs here in the paralogisms. But freedom is not mentioned at all in this context. It is instead one of the four cosmological problems described in the antinomies. Nevertheless, the paralogisms are helpful for understanding the nature of the antinomies. Each species of dialectical inference follows the same pattern. It takes a legitimate transcendental idea (the soul, the world, God), but it then makes the mistake of conflating a subjective condition of thought for a condition of a given object. We then erroneously conclude that we can have objective knowledge of

---

use is precluded by the fact that the alleged object to which it is being applied is not empirical. Even more problematically, on Kant's view, there is *no object given at all*" ("Kant's Critique of Metaphysics").

[88] The fourth paralogism is a bit of an outlier. In the second edition of the *Critique* Kant revised this section substantially, but the most notable change was that the fourth paralogism was recast entirely as the "Refutation of Idealism" and was moved to the Analytic.

something which lies beyond the boundaries of experience as if we were cognizing a determinate object of experience.

The difference with the antinomies is that the paralogism "effects a merely one-sided illusion regarding the idea of the subject of thought" (A407 / B434). When we contemplate the nature of the thinking subject, we are tempted only in one direction. Kant thinks that we are naturally (though mistakenly) drawn to make conclusions like those of Leibniz and Descartes.[89] It looks as if the soul is a simple substance; it persists through time; etc. Although the antinomies come about through a similar process (*viz*., thinking about a transcendental idea qua unconditioned condition), the result is notably distinct. In the context of thinking about our transcendental idea of the world, reason is led "into the temptation either to surrender itself to a skeptical hopelessness or else to assume an attitude of dogmatic stubbornness, setting its mind rigidly to certain assertions without giving a fair hearing to the grounds for the opposite" (A408 / B434). Kant thinks that we are pulled in two directions. We find the skeptical conclusions of "empiricism" on one side and "dogmatic rationalism"[90] on the other.

In spite of this point of contrast with the paralogisms, the antinomies are generated in a similar fashion. We misuse a transcendental idea in such a way that we come to unjustified conclusions about it. But when it comes to our idea of the world as a totality, we arrive at seemingly

[89] Kant lends no credence whatsoever to Hume's skeptical conclusion in the *Treatise* that we must reject the idea of a thinking self on the grounds that we cannot trace it back to any impression. Kant says that "there is not the least plausibility" for such a claim (A406 / B433).

[90] These labels make it sound like Kant has in mind only the British empiricists like Locke, Berkeley, and Hume and the continental rationalists like Descartes and Leibniz. But Kant thinks of himself as describing a much older conflict, placing Epicureans on one side and Platonists on the other. He is very clear about framing in those historical terms: "This is the opposition of **Epicureanism** and **Platonism**" (A471 / B499). Nevertheless, we should not downplay the importance of the debates that were going on in Kant's time. Most notably the antinomies very closely reflect some of the debates in the Leibniz-Clarke correspondence. Michelle Grier says, "the thesis positions of the antinomies, particularly the two mathematical antinomies, clearly map onto the Newtonian views as articulated by Clarke, whereas the Antithesis positions reflect Leibniz's responses to Clarke" (*Transcendental Illusion*, 182).

contradictory conclusions. The first antinomy, for instance, involves the question of whether or not the world has a beginning in time.[91] If we assume that the world has no beginning in time, then "up to every given point in time an eternity has elapsed, and hence an infinite series of states of things in the world, each following another, has passed away" (A426 / B454). But this presupposes that an infinite series has come to completion "through a successive synthesis." This is absurd, however; Kant thinks it is akin to counting to the highest integer. On the other hand, if we assume that the universe does have a beginning in time, we are no better off. If the universe had a beginning in time, then "there must be a preceding time in which the world was not, i.e., empty time." Kant thinks this is equally absurd. It is impossible for the world (and time) to come into existence, since the very idea of coming-to-be presupposes time. As Kant puts it, "no arising of any sort of thing is possible in an empty time" (A427 / B455).

Reason finds itself "embarrassed"[92] by this conflict. Although we could potentially be content with the conclusions of the paralogisms (if we failed to see the source of error), we cannot be satisfied when confronted with an antinomy.[93] Kant says that leaving such a contradiction unresolved would be "the death of a healthy philosophy," — the "**euthanasia** of pure reason" (A407 /B434). Fortunately for the fate of reason, the antinomy can be resolved. The key to

[91] It also asks whether or not it has a limit in space, but for the sake of simplicity, I have restricted my discussion to the first half of the antinomy.

[92] This is Kant's wording. In his discussion of the third antinomy, he says, "the question of the freedom of the will" has "always put speculative reason into such embarrassment" (A448 /B476).

[93] It is widely known that Kant says in the *Prolegomena* that Hume's skeptical doubts awakened him from his "dogmatic slumber" and prompted him to devise the apparatus of the first *Critique*. Far fewer are likely to have read a letter that Kant wrote to Christian Garve in 1798 in which he claimed that it was the *antinomies* that first aroused him from his dogmatic slumber! He writes, "It was not the investigation of the existence of God, immortality, and so on, but rather the antinomy of pure reason — "The world has a beginning; it has no beginning, and so on, right up to the 4th [*sic*]: There is freedom in man, vs. there is no freedom, only the necessity of nature" — that is what first aroused me from my dogmatic slumber and drove me to the critique of reason itself, in order to resolve the scandal of ostensible contradiction of reason with itself" (Kant to Garve, 21 September 1798; 12: 258). Of course, there is not necessarily any inconsistency between these two descriptions of his awakening. His reflections on Hume's empiricism might be what prompted him to discover the antitheses.

eradicating the contradiction lies in understanding how reason's demand for the unconditioned is being misused.

The proof of the thesis and the antithesis both take the form of a reductio. The thesis assumes that there is no beginning in time and discovers an absurdity; the antithesis supposes that there is a beginning in time and similarly derives a contradiction. The legitimacy of this method depends on the following disjunction: there is a world, and it is either finite or infinite. Kant's solution is to reject the disjunction altogether. The world is neither infinite nor finite. More precisely, we are predicating 'finite' or 'infinite' of an illegitimately hypostatized subject — 'the world.'[94] When we treat our idea of the world as the subject of predication, we inevitably end up equivocating. If we are talking about a set of spatiotemporal appearances, then it must go back infinitely in time. Appearances are necessarily temporal, and this proscribes the possibility that time has an ultimate beginning. If, on the other hand, what we mean by 'the world' is a transcendental idea of pure reason, then it must have a beginning in time, since reason refuses to permit an infinite regress. Thus, the reductio of the thesis fails because it has *not* successfully demonstrated that the world (qua set of appearances) must have a beginning in time; it has shown that our idea of a temporal series requires a starting point. The reductio of the antithesis fails in the very same way. The antithesis has shown merely that our idea of pure reason (the world qua unconditioned totality) is inconsistent with empirical conditions. Either way, we are trying to hold an idea accountable to conditions that do not belong to it. For the proof of the thesis, we try to hold appearances accountable to our idea of an unconditioned totality. In the antithesis, we try to find a home for our pure idea in an empirical world whose conditions exclude its possibility. But the

[94] The same goes for the thinking subject. When it comes to the pure 'I' of apperception, we can say very little, since we cannot cognize the subject. We must posit an 'I' which synthesizes representations, but we cannot have any cognition of it. Kant says, "I therefore have **no cognition** of myself **as I am**, but only as I **appear** to myself" (B158).

predicates ‘finite’ and ‘infinite’ have no place here, for the subject — ‘the world’ — is not given as an object of experience to which the disjunction could apply.

In this way, we can see how the dialectical inferences of the antinomies are similar to the paralogisms. Kant claims that the arguments of the antinomies also rely on the fallacy of equivocation. We begin with reason’s logical demand that we seek an unconditioned condition. But, as we saw, this logical principle then takes a metaphysical form: “If the conditioned is given, then the whole series of all conditions for it is also given” (A409 /B436). This is the major premise of the syllogism. The minor premise is that “objects of the senses are given as conditioned” (A497 / B525). The conclusion from this is that the whole series of conditions is given for objects of the senses. This argument commits the fallacy of equivocation /ambiguous middle as it equivocates on the term ‘conditioned.’ Kant says, “it is clear that the major premise of the cosmological syllogism takes the conditioned in the transcendental signification of a pure category, while the minor premise takes it in the empirical signification of a concept of the understanding applied to mere appearances; consequently there is present in it that dialectical deception that is called a ‘*sophisma figurae dictionis*’” (A499–500 / B527–28). Michelle Grier gives a lucid explanation of the fallacy in this context:

> What Kant means here is that the major premise uses the term “the conditioned” in a very general way, one that considers things in abstraction from the sensible conditions of our intuition. The minor premise, however, which specifically refers to objects in space and time (appearances), is committed to an empirical use of the term. Indeed, such an empirical use would have to be deployed, if the conclusion is to be reached. The conclusion is that the entire series of all conditions of appearances is *actually given*. Put in other terms, the conclusion is that *there is a world*, understood as the sum total of all appearances and their conditions (A420 /B448). This hypostatization of the idea of the world, the fact that it is taken to be a mind-independent object, acts as the underlying assumption motivating both parties to the two mathematical antinomies.[95]

[95] Grier, “Kant’s Critique of Metaphysics.”

What happens in the first two (mathematical) antinomies is a conflation of the legitimate and illegitimate applications of reason's demand for the unconditioned, and this leads us to two false conclusions. We come to think that the world (as an unconditioned totality) is *given* as a condition of appearances, and then we are led to mistaken beliefs about it.

The third and fourth antinomies (which Kant refers to as the "dynamical antinomies") do not share this feature. The dynamical antinomies, like the mathematical ones, are brought about by reason's principle that an unconditioned condition must be given for a conditioned series. But unlike the mathematical antinomies, when it comes to the dynamical antinomies, Kant says that the thesis and antithesis "may **both be true**" (A532 / B560). The reason for this difference is that the mathematical antinomies concern only the world of spatiotemporal objects. The first antinomy asks whether or not the world has a beginning in time. The domain is strictly limited to the temporal series. The debate is whether or not the temporal series has a first member or whether it goes back infinitely into the past. The same is true of the second antinomy, which concerns the infinite divisibility of objects. Both of these concern the spatiotemporal world and out attempt to hold it accountable to the subjective conditions of thought.[96]

[96] In his discussion of the resolution to the mathematical antinomies, Kant offers the following comparison: "If someone said that every body either smells good or smells not good, then there is a third possibility, namely that a body has no smell (aroma) at all, and thus both conflicting propositions can be false" (A503 / B531). So it is with the question of whether the world is finite or infinite. The predicates 'finite' and 'infinite' cannot properly be ascribed to the subject, 'the world.' So rather than comparing it to an object that doesn't have a smell (like distilled water), it might be better to compare it to asking whether the number seven smells good or bad. The disjunction must be rejected as a category mistake; the number seven smells neither good nor bad.

The dynamical antinomies have a different form. We can compare it to looking at a spoon. On one side, someone claims that the object is concave. On the other side, someone claims that the object is convex. But, as a transcendental idealist, we have a perspective that is available to neither of the disputants. We can say that both claims are true. The object is concave *and* convex. The seemingly contradictory positions turn out to be consistent with one another when we adopt transcendental idealism.

2.3.3 The Third Antinomy

It is the Third Antinomy that deals with the problem of freedom and determinism. The thesis holds that there must be a free cause, and the antithesis argues that everything happens solely in accordance the deterministic laws of nature. But in this case, unlike the mathematical antinomies, it is possible for both the thesis and the antithesis to be true, since the free cause is found in the noumenal world, and the laws of nature apply only to the phenomenal world. Indeed, Kant thinks reason cannot forfeit on either side. In this section, I reconstruct those arguments and explain how Kant resolves the conflict between them.

The thesis is the following claim: "Causality in accordance with laws of nature is not the only one from which all the appearances of the world can be derived. It is also necessary to assume another causality through freedom in order to explain them" (A442 / B472). The proof for the thesis (like all the others) takes the form of a reductio and it goes as follows:[97]

> (P1) Assume (for reductio) "that there is no other causality than that in accordance with laws of nature."
>
> (P2) If (1), "then everything that happens presupposes a previous state, upon which it follows without exception according to a rule."[98]
>
> (C1) "Everything that happens presupposes a previous state, upon which it follows without exception according to a rule."
>
> (P3) The previous state "must be something that has happened."
>
> (C2) Every event presupposes a previous event, which in turn presupposes a previous event.[99]
>
> (P4) There is never a first cause "and thus no completeness of the series on the side of the causes descending one from another."

---

[97] My reconstruction follows Kant's text as closely as possible. It is also indebted to Eric Watkins. See *Kant's Metaphysics of Causality*, 306. Cf. Michelle Grier, *Kant's Doctrine of Transcendental Illusion*, 214.

[98] By 'according to a rule,' Kant means a law of nature.

[99] As Kant puts it, "the causality of the cause through which something happens is always something **that has happened**, which according to the law of nature presupposes once again a previous state and its causality, and this in the same way a still earlier state, and so on."

(P5) If (P4), then there is something that happens "without a cause sufficiently determined a priori."

(C3) Something happens "without a cause sufficiently determined a priori."

(P6) "Nothing happens without a cause sufficiently determined a priori."

Therefore, P1 is false (A444–45 / B472–73).

Arguably, the most controversial premise in the argument is P5.[100] According to that premise, if there is an infinite series of causes, then there is something that happens without a cause sufficiently determined a priori; that is, an infinite series of causes violates the principle of sufficient reason.[101]

Kant's conclusion here is that "the law of nature, when taken in its unlimited universality, contradicts itself, and therefore this causality cannot be assumed to be the only one" (A446 / B474). The law of nature requires that everything has an explanation a priori, but he thinks that this leads us to consider the entire series of causes. That series (considered as a whole) does *not* have an explanation. Kant thinks that there must be a completed series for there to be a satisfying causal explanation. The fifth premise says that if there is an infinite chain of causes going back into the past, then there is something that happens without a *ratio*. He is employing a very strong version of the principle of sufficient reason here. According to this principle, a sufficient explanation of

[100] As Watkins points out, P2 is an analytic truth for Kant. If we assume that everything happens according to the laws of nature, then every event is determined by a preceding state and a rule. P3 might seem somewhat controversial. Kant's proof for it is that no event could have followed from a state which existed eternally. But if we keep in mind that Kant is talking about an event in time, then this premise should not trouble us. Events in time are caused by other events in time, not by eternally existing states.

[101] This might come as a surprise to a Spinozist, who is committed to both the principle of sufficient reason and the idea that the chain of causes goes back infinitely into the past. But Kant would point out that the Spinozist already has an answer to the pressing question. Kant's worry is that the entire series of causes stands in need of a cause. It would not be a transitive cause like the members of the causal series, but it must be something that explains the existence of the series of causes. For the Spinozist, as I explained above, this would be God as the immanent cause.

event requires two things. First, there must be a preceding state from which it followed according to a rule. Second, there must be a complete explanation in the sense that there is "nothing further to be explained."[102] If the series of causes goes back infinitely into the past, then the explanation is never complete. There would always be another event to be explained.[103] Thus, Kant thinks we are compelled to posit a first cause. The principle of explanation which sent us out in search of a natural cause surprisingly led us to the conclusion that there must be something other than just the causality of nature.

But reason is drawn equally in the opposite direction. The antithesis holds that "There is no freedom, but everything in the world happens solely in accordance with the laws of nature" (A445 / B473). The proof for this antithesis adopts the same approach as the thesis. It asks us to assume for reductio that transcendental freedom is a first cause. After looking closely at the conditions of transcendental freedom, we realize that it leads to a contradiction and must be rejected. Recall that one of the conditions of transcendental freedom is that it begins a series *spontaneously*, and, for Kant, this means that "nothing precedes it through which this occurring action is determined in accordance with constant laws" (A445 / B473). The problem with this is that the idea of something happening or causing something to happen *necessarily* involves a preceding state. This was a crucial step in the argument of the thesis: "everything **that happens**

[102] Allison, *Transcendental Idealism*, 380.

[103] Bennett complains that this version of the PSR is much too strong and that it turns Kant's argument into a straw man. Allison retorts that Leibniz holds this very position. He says, "Leibniz maintains that every occurrence has a sufficient reason *both* in the sense that it has an antecedent cause *and* in the sense that it has an ultimate explanation (accessible only to God) based on its role within the total context of the possible world actualized by the divine will" (*Transcendental Idealism*, 380).

For Leibniz, there would be more than one answer to the question of why a particular event happened. First, there would be a causal account about a substance and its antecedent states. But there would also be a story about why these substances exist rather than others, why these antecedent states came into being, etc. The answer to those questions would be that God chose to create the best of all possible worlds. Thus the strong version of the PSR would be satisfied. It ends with an "ultimate explanation." Kant is looking for something similar. He thinks that reason demands an explanation of the entire causal history.

presupposes a previous state, upon which it follows without exception according to a rule" (A444 / B472). The absolute spontaneity of transcendental freedom contradicts "the causal law." Positing transcendental freedom might satisfy reason's demand for the unconditioned, but it comes at an unacceptable cost. All of our scientific knowledge is premised on the causal principle. In science, we are seeking the rules that explain why events followed from preceding states. And transcendental freedom tells us that some events did not follow from rules at all; they began spontaneously. "Thus nature and transcendental freedom are as different as lawfulness and lawlessness" (A447 / B475). Ironically, reason's quest for unity and completeness ended up creating a grave threat to the unity of nature.

Reason is torn. It faces a problem generated by its own project of seeking a unified explanation of the world. Our best (and indeed only) principle for explaining events in the world is the causal principle: we know a priori that every event is determined by a preceding state according to a rule. In its search for a *complete* explanation of the world, however, reason finds itself compelled to posit a first cause — one that is truly spontaneous. But this supposition is at odds with the scientific principle which confers unity on our experience and makes knowledge of the world possible. And unlike the resolution to the mathematical antinomies, we cannot deny this disjunction. Although it was possible for us to affirm that the world is neither infinite nor finite, we cannot reject both freedom and nature. After all, one of the principal aims of the *Critique of Pure Reason* was to provide a firm foundation for science. If Kant were to reject his causal principle, then science would be doomed.[104] But a complete explanation of the world requires

[104]More precisely, science *as Kant understands it* would be doomed. After rejecting Kant's causal principle, science might very well carry on in a Humean fashion. We would deny the claim that causes necessitate their effects, and we would just make probability claims. As a matter of custom or habit, we would know that the presence of certain "causes" means that certain "effects" are likely to follow. But this view of nature and science was unacceptable to Kant. Long before the debate over quantum mechanics, Kant held something akin to Einstein's position that "God

freedom. And, as I explain in the next chapter, without freedom Kant's moral theory would be in serious jeopardy. Hence the resolution of the antinomy must explain how it is possible for both the thesis and the antithesis to be true.

2.4 The Resolution of Reason's Conflict with Itself

The key to the resolving the antinomy is understanding how the contradiction arises only if one fails to properly distinguish between things in themselves and appearances. Kant calls this view 'transcendental realism,' and he argues that the conflict between **nature** and **freedom** is dissolved once we abandon this view and replace it with transcendental idealism. As we saw in chapter 1, transcendental freedom was defined as an absolute causal spontaneity — the capacity to act as a first cause without any causal antecedent. This requires complete independence from the laws of nature and the empirical world. Obviously, this concept cannot find a home in the phenomenal world, which is necessarily governed by natural laws. Thus, Kant must attribute transcendental freedom to things in themselves, not appearances. He argues that rational beings like us could possess transcendental freedom as intelligible beings and that our free, noumenal choices could have effects in the world of appearance. In this section, I explain the details of this view, and I conclude by underscoring its limited scope. After showing that freedom and nature are not incompatible with one another, Kant emphasizes that this is not enough for us to know the reality or even the metaphysical possibility of transcendental freedom. The reader must take the entire resolution with a grain of salt. Kant has not offered a view of how freedom *actually* works;

---

does not play dice with the universe." For Kant, if you were to have knowledge of the laws and of the initial conditions, then, like Laplace's demon, you would be able to predict every event with perfect precision, as the effects would be necessitated by their causes according to the deterministic laws of nature. Kant was not willing to accept Hume's view that science should be the mere practice of getting better at anticipating probable events. Philip Kitcher's "Projecting the Order of Nature" contains an excellent discussion of how Kant's view provides an alternative to both the Humean view and the view that our scientific knowledge "carves nature at the joints" by describing mind-independent facts about nature and causality.

he has only given a sketch of how it *might* work. All he wants to do is demonstrate that freedom is not logically impossible and that it is compatible with the causal principle of the phenomenal world. To draw any conclusion stronger than that would be to engage in unjustified metaphysical speculation.

2.4.1 The Compatibility of Nature and Freedom

Kant frames the issue in terms of a conflict between nature and freedom, and he gives diametrically opposed definitions for them:

> In respect of what happens, one can think of causality in only two ways: either according to **nature** or from **freedom**. The first is the connection of a state with a preceding one in the world of sense upon which that state follows according to a rule . . . By freedom in the cosmological sense, on the contrary, I understand the faculty of beginning a state **from itself**, the causality of which does not in turn stand under another cause determining it in time in accordance with the law of nature. (A532–33 / B560–61)

Defined in this way, the two terms are thoroughly incompatible with one another. And since the "correctness of the principle of the thoroughgoing connection of all occurrences in the world of sense according to invariable natural laws is already confirmed as a principle of the transcendental analytic" the only question that remains is whether or not there is any room left for freedom. We must ask "whether freedom is possible anywhere at all, and if it is, whether it can exist together with the universality of the natural law of causality, hence whether it is a correct disjunctive proposition that every effect in the world must arise **either** from nature **or** freedom, or whether instead **both**, each in a different relation, might be able to take place simultaneously in one and the same occurrence" (A536 / B564).

Fortunately, transcendental idealism makes room for this possibility. Although freedom and nature have been given inconsistent definitions, it is possible for them to coexist as long as

they have separate domains: freedom has its home in the noumenal world and the laws of nature belong only to the world of phenomena.[105] As such, the antinomy constitutes additional evidence for transcendental idealism, since "the common but deceptive presupposition of the absolute reality of appearance immediately shows its disadvantageous influence for confusing reason. For if appearances are things in themselves, then freedom cannot be saved" (A536 / B564). In his initial description of the solution, Kant says that transcendental idealism allows us to see how libertarian freedom might be compatible with deterministic laws of nature:

> 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. Thus the intelligible cause, with its causality, is outside the series; its effects, on the contrary, are encountered in the series of empirical conditions. The 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; this is a distinction which, if it is presented in general and entirely abstractly, must appear extremely subtle and obscure, but in its application it will be enlightening. (A536–37 / B564–65)

He is surely right that the distinction is "subtle and obscure," for he is claiming that the deterministic laws of nature are somehow *not* compromised by the fact that spontaneous, uncaused causes from the noumenal world have tangible effects in the phenomenal world. How is this possible? When we consider a voluntary human action, there are now two causal stories to tell. I got up to make myself a cup of tea, and there are two accounts of what caused the event. On the one hand, this event was completely determined by empirical laws of nature. These laws were in place before I ever existed, and it seems that I have no control over them. So it looks like the event

[105] Allen Wood quips that this is like saying that a married couple is compatible "only as long as they live in separate houses." See "Kant's Compatibilism," 241. Of course, the two worlds are not completely isolated from one another. Kant repeatedly states that noumenal choices have *effects* in the phenomenal world. Thus, he will owe us an account of how noumenal choices might manifest themselves in a world governed by natural laws.

is completely outside of my control. On the other hand, I exist as a noumenal being, and my free choices as a noumenal being could be "intelligible causes" have "effects" in the phenomenal world. My getting up to make tea might be one of these effects.

It would seem that a third story is needed here; namely, we need an account of how these two causal explanations are compatible with one another. So far we have been told that, as an event in the phenomenal world, my getting up to make tea was determined by laws of nature. And, as a noumenal being, I make free choices that have effects in the phenomenal world. One way to explain how noumenal choices might have effects in the phenomenal world would be to claim that noumenal choices interrupt the chain of empirical causes and break the laws of nature. On this account, *most* events in the world would be determined by natural laws, but the noumenal self would occasionally make miraculous interventions by defying the natural laws. Some events in the world would lack deterministic causes.[106]

Of course, this kind of "miraculous intervention" account is not an option for Kant. He is adamant that *every* event in the natural world is necessitated by laws of nature. Without that principle, knowledge of nature would be impossible. There can be no exceptions the laws of nature. My getting up to make tea must have a naturalistic causal explanation — one that explains how my action was necessitated by inviolable laws. Insofar as we are talking about an action in space and time, this event can and should be explicable in the same way that we would explain a planet's

[106] In the contemporary free will literature, such incompatibilist views are referred to as "noncausal" theories. Some proponents of noncausal theories argue that free actions must not be caused *by anything at all*; others argue that free actions might be caused, as long as their causes were not deterministic. See Carl Ginet, "Reasons Explanations of Action: Causalist versus Noncausalist Accounts." Cf. Hugh McCann, *The Works of Agency: On Human Action, Will, and Freedom*.

To some extent, Kant might be seen as an ally of the noncausal view. When it comes to transcendental freedom, an action must be truly uncaused. But Kant can only support these views to a limited extent, since many proponents of these views limit their scope to the phenomenal world. They view certain events in the phenomenal world as uncaused (or as caused in some indeterministic way), and Kant cannot allow that. But Kant does share with them the view that genuine spontaneity requires that the free choice be completely exempt from causal determination.

rotation around the sun. Kant's account must tell us how noumenal choices could manifest themselves in the empirical world *without* recourse to miraculous interruptions.

Kant recognizes this constraint, and he accordingly offers an account that reconciles nature and freedom in such a way that the latter does not compromise the integrity of the former. For reasons that will become clear in chapter 4, Kant cannot commit himself to any particular account of exactly how this works. But the resolution of the antinomy requires some explanation of how it *might* work. In brief, the view he suggests is that our noumenal selves ground not only the existence of phenomenal appearances; they also ground the *laws* that govern those appearances.[107] He puts this most plainly in the *Groundwork* where he says that "the world of understanding [i.e., the noumenal world] contains the **ground** of the world of sense [the phenomenal world] [108] and **so too of its laws**" (4:453; emphasis added). The laws of nature operate on their own and have no exceptions or interruptions. But if we were to push our inquiry further and ask why these laws of nature exist rather than others, the answer could be that the free choices of noumenal selves ground laws of the empirical world.

This is still much too vague however; Kant says at this point that he has only begun to "sketch the silhouette of a solution" (A542 / B570). He provides a number of important details in order to flesh out the account. At the heart of the issue is the idea of our "empirical character." What Kant means by 'character' in this context is that every being in the empirical world has dispositions such that it necessarily follows natural laws. As he puts it:

---

[107] This is the interpretation favored by Eric Watkins in *Kant and the Metaphysics of Causality*. It is the one that I lean toward, but as I explain, it is only one of several possibilities, and we should remain noncommittal.

[108] In this passage (and throughout the *Groundwork*), Kant refers to the noumenal realm as 'the world of understanding' [*Verstandeswelt*] and the phenomenal world as 'the world of sense' [*Sinnenwelt*]. Kant uses a variety of words for the noumenal self and its realm. In the first *Critique* he talks mostly about "intelligible causes" and the "intelligible character" which grounds the "empirical character." All of these terms (intelligible self, noumenal self, self as it is in itself, member of the world of understanding) refer to the same thing, and so I use them interchangeably.

> [E]very effective cause must have a **character**, i.e., a law of its causality, without which it would not be a cause at all. And then for a subject of the world of sense we would have first an **empirical character**, through which its actions, as appearances, would stand through and through in connection with other appearances in accordance with constant natural laws, from which, as their conditions, they could be derived; and thus, in combination with these other appearances, they would constitute members of a single series of the natural order. (A539 / B567).

We can think of our empirical character as a set of psychological dispositions. If we were to have perfect knowledge of someone's empirical character *and* we knew all of the psychological /anthropological laws that govern human behavior, then we could predict every action with perfect certainty.[109] In principle, it would be no different than plugging numbers in for the variables in Kepler's equations for planetary motion.

But in addition to empirical character, human beings also have an intelligible character:

> [O]ne would also have to allow this subject an **intelligible character**, through which it is indeed the cause of those actions as appearances, but which does not stand under any conditions of sensibility and is not itself appearance. The first one could call the character of such a thing in appearance, the second its character as a thing in itself. (A539 / B567).

This intelligible character is a property of a thing in itself (*viz.*, the noumenal self), and the empirical character is a property of an appearance (the phenomenal self, the self as it appears). And given that the noumenal self is timeless, it is exempt from causal necessity:

> [I]n its intelligible character (even though we can have nothing more than merely the general concept of it), this subject would nevertheless have to be declared free of all influences of sensibility and determination by appearances; and since, in it, insofar as it is a **noumenon**, nothing **happens**, thus no alteration requiring a dynamical time-determination is demanded, and hence no connection with appearances as causes is encountered in its actions, this active being would to this

[109] Kant's affirmation of this idea is unequivocal: "Because this empirical character itself must be drawn from appearances as effect, and from the rule which experience provides, all the actions of the human being in appearance are determined in accord with the order of nature by his empirical character and other cooperating causes; and if we could investigate all the appearances of his power of choice down to their basis, then *there would be no human action that we could not predict with certainty, and recognize as necessary given its preceding conditions*" (A549–550 / B577–78; emphasis added).

> extent be independent and free of all the natural necessity present solely in the world of sense (A540–41 / B568–69)

The noumenal status of the intelligible self makes room for this exemption, but it also places a restriction on our knowledge of it. Kant holds that we do not have any theoretical cognition of things in themselves, and thus we cannot know the noumenal self in the robust, substantive way that we know about objects of experience:

> This intelligible character could, of course, *never be known immediately*, because we cannot perceive anything except insofar as it appears, but it would have to be **thought** in conformity with the empirical character, just as in general we must ground appearances in thought through a transcendental object, even though *we know nothing about it as it is in itself.* (A540 / B568; italics added).

Just like every other thing in itself, we must admit a certain kind of ignorance when it comes to the nature of the noumenal self and its intelligible character. We cannot know what it is like, and we cannot have a fully determinate cognition of it. But there is at least one important exception to the principle of noumenal ignorance.[110] One thing that we know about things in themselves is that they ground appearances.[111] Thus, our empirical character is grounded by our intelligible character.

This dual character of the human being is what allows Kant to claim that libertarian freedom does not compromise the deterministic laws of nature. In the phenomenal world, actions are "always determined beforehand by empirical conditions in the preceding time, but only by

---

[110] Depending on how the principle is defined, there might be *several* exceptions. For instance, we know that things in themselves are not in space and time; we know that they are not subject to the laws of nature; etc. In order to make the principle of noumenal ignorance exceptionless, it would have to be amended. We could change it to say something like "we have no *substantive* knowledge of things in themselves," and 'substantive knowledge' would have to denote something like 'determinate cognition.' Even that statement might be too strong, however. It will become clear in chapter 3 that practical knowledge might yield some cognition pertaining to things in themselves. This would call for a further addendum, leaving us with the following: "we can have no particular, substantive, theoretical knowledge of things in themselves." The principle of noumenal ignorance is exceptionless only if it is qualified in this way.

[111] In one of his many statements of the "grounding thesis" (Watkins's term), Kant writes "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; boldface added). Kant is very clear (here and elsewhere) that things in themselves ground appearances.

means of the empirical character (which is a mere appearance of the intelligible character), and they are possible only as a continuation of the series of natural causes. Thus freedom and nature, each in its full significance, would both be found in the same actions, simultaneously and without any contradiction, according to whether one compares them with their intelligible or their sensible cause" (A541 / B 569). Every empirical question admits of an empirical answer. All human actions in the world of space and time are causally necessitated by a law in accordance with that person's empirical character. But their empirical character is not their *only* character. They also have an intelligible character that grounds the empirical.

In order to make the idea easier to understand, consider the following analogy.[112] A neuroscientist has created a machine that allows people to choose a plotline for a dream that they would like to have. The dream will be vivid and lifelike; the experiences one has in the dream machine will be nearly indistinguishable from real experiences. One day, Arnold decides to visit the scientist in order to have an experience that he cannot have in the real world. He wants to go snorkeling in the Great Barrier Reef, and he wants to interact with sharks. Arnold tells the scientist what he wants to see and do in his dream, and the scientist programs the machine accordingly. Once he enters the dream, all of his actions will be dictated by the machine's programming. The scientist can predict with perfect certainty everything that will happen in Arnold's dream: Dream-

[112] This analogy is my own, but it is a combination of Derk Pereboom's lucid dream analogy and a bit of science fiction. See Pereboom, "Kant on Transcendental Freedom." Pereboom's example involves him having a lucid dream in which he is capable of freely deciding what will happen in the dream. He chooses to have a dream in which a group of Martian neuroscientists use mind control which compels him to steal the Mona Lisa. Thus, there is a sense in which he freely chose to steal the painting (he chose to have this dream), but there is also a sense in which it was causally necessitated (by the Martians' mind control).

My example is similar, but it has the important difference that the existence of the deterministic component is not freely willed by the dreamer. Instead, the deterministic algorithm of the dream machine is a necessary condition of the existence of the dream world. In Pereboom's lucid dream, the dreamer wills the existence of determinism. Pereboom's lucid dream could very well carry on without the Martians and their deterministic mind control. The difference is subtle, but it makes my example a slightly better fit with Kant's account. Although the particularities of one's empirical character (and which laws govern it) might be a consequence of what is freely willed by the noumenal self, the *existence* of deterministic constraints is not a product of free will. In my example, the deterministic algorithm is a condition of the possibility of the dream world rather than a freely willed component of the dream.

Arnold has no freedom whatsoever. In the dream, Arnold has the thrilling experience of reaching out to grab the dorsal fin of a great white shark. Weeks later, after the dream is over, Arnold asks himself the question, "What caused me to grab the shark's fin?"

As he wrestles with this question, Arnold finds himself pulled in different directions. There are two reasonable causal explanations, depending on the scope of the question. One answer is that all of Dream-Arnold's actions happen in accordance with the dream-machine's programming. The character of Dream-Arnold follows the parameters entered by the neuroscientist. Dream-Arnold is adventurous; that's why he was bold enough to reach out and grab the shark. Dream-Arnold does precisely what his dream-character determines him to do. The second answer is that Real-Arnold told the scientist to program the machine such that he would have this experience. The character and decisions of Real-Arnold are the ground for the character and decisions of Dream-Arnold. Real-Arnold made free choices that gave rise to deterministic algorithms, which completely determined the behavior of Dream-Arnold.

Aside from a handful of disanalogies,[113] this is an illuminating thought experiment. It allows us to conceive of a single event with two distinct causal explanations (one free and one determined), *and* it avoids the problem of having free choices interrupt the series of deterministic causes. Within the context of the dream, every action of Dream-Arnold's is determined by exceptionless laws. Everything that happens in the dream was predetermined by the machine's programming. But the rigid determinism within the dream does not have any bearing on Real-

[113] One important difference is that Real-Arnold stands in a temporal relation to Dream-Arnold. Real-Arnold's decisions causally condition Dream-Arnold's experience in a way that is importantly distinct from the intelligible character's grounding relationship with empirical character. The intelligible self is timeless; so it cannot stand in a temporal relation to the empirical self. Furthermore, the dream analogy has the problem that it makes the empirical world sound like it has the ontological status of a dream or illusion. Kant does not intend for us to think of the empirical world as a mere fantasy. But this difference is not terribly troubling. Although the spatiotemporal world is not a dream, it *is* nevertheless dependent on thinking beings. It is a world of representation. Kant's transcendental idealism is, after all, *idealism* even if it's not of the Berkeleyian variety.

Arnold's freedom when he told the scientist that he wanted to have the experience of grabbing a shark's dorsal fin. Dream-Arnold had no choice in the matter, but the free decisions of Real-Arnold are in no way compromised by determinism in the dream world.

Once we have a handle on this idea of the human being's dual character, we can understand what Kant means when he says the "causality of an empirical cause" might not be empirical. Although every empirical event has an empirical cause, it is possible that there is a free, intelligible cause behind the entire causal series:

> But then if the effects are appearances, is it also necessary that the causality of their cause, which (namely, the cause) is also appearance, must be solely empirical? Is it not rather possible that although for every effect in appearance there is required a connection with its cause in accordance 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*, i.e., an original action of a cause in regard to appearances, which to that extent is not appearance but in accordance with this faculty intelligible, even though otherwise, as a link in the chain of nature, it must be counted entirely as belonging to the world of sense? (A544 / B572; emphasis added).

Although every empirical event has a deterministic empirical cause, this does not preclude the possibility that a noumenal being might freely choose its character, and this intelligible character could have effects in the empirical world, as it manifests itself (or appears) as the subject's empirical character.

In spite of his usual reluctance to provide concrete examples, Kant demonstrates how this idea of the human being's dual character could be applied in the real world. This was the case of the "malicious lie" discussed in the previous chapter. When dealing with a "voluntary action" of this sort, we are drawn to two different modes of explanation. First, we might consider the action empirically. Insofar as we are asking an empirical question about what caused the malicious lie, we are seeking a strictly naturalistic explanation of antecedent causes; we might go looking for

history of "a bad upbringing, bad company" or other factors that led to the liar's wicked "natural temper" (A554 / B582). But even after concluding that the action can be fully explained by these empirical causes, which seem to be entirely outside the agent's control, "one nonetheless blames the agent" (A555 / B583). Kant thinks that this moral blame requires us to "entirely set aside how that life was constituted, and consider that the entire series of conditions that transpired might not have been." This is possible for us only because we can imagine that empirical world would have been different if the liar were to have a different intelligible character. Since "the action is ascribed to the agent's intelligible character: now in this moment when he lies, it is entirely his fault" (A555 / B583). In the dream world, Arnold had no choice but to grab the shark's fin, but if Real-Arnold had chosen to have a different dream, then the entire series of events leading up to grabbing the fin would never have taken place. So it is with the malicious lie. As a consequence of antecedent causes, the lie was causally necessitated and unavoidable; if, however, we turn our focus to the agent's intelligible character, things might have (and should have) turned out differently.

2.4.2 The Epistemic Status of Transcendental Freedom

At this juncture, it would be reasonable to ask about the epistemic status of this account. Are we capable of knowing that we possess this kind of freedom? Does this account demonstrate that transcendental freedom is metaphysically possible? Kant's answer to both questions is a resounding 'no':

> It should be noted that here *we have not been trying to establish the reality of freedom*, as a faculty that contains the causes of appearance in our world of sense. For apart from the fact that this would not have been any sort of transcendental investigation having to do merely with concepts, it could not have succeeded, since from experience we can never infer something that does not have to be thought in accord with the laws of experience. Further, *we have not even tried to prove the* **possibility** *of freedom*; for this would not have succeeded either, because from mere concepts a priori we cannot cognize anything about the possibility of any real

> ground or any causality. *Freedom is treated here only as a transcendental idea*, through which reason thinks of the series of conditions in appearance starting absolutely through what is sensibly unconditioned, but thereby involves itself in an antinomy following its own laws, which it prescribes for the empirical use of the understanding. [To show] *that this antinomy rests on a mere illusion, and that nature at least* **does not conflict with** *causality through freedom* — that was the one single thing we could accomplish, and it alone was our sole concern. (A557–58 / B585–86; italics added)

This is Kant's final remark on the Third Antinomy as he transitions after this to his discussion of a necessary being. Several points are worth noting here. First, Kant denies that he has established the reality of freedom, and he does so for two reasons. It would be impossible to demonstrate the reality or existence of something from mere concepts.[114] Furthermore, the idea in question — transcendental freedom — is not "thought in accord with the laws of possible experience," and no experience of something conditioned can justify an assertion of the unconditioned, as that would extend our cognition beyond the boundaries of experience. Second, Kant says that he has not even attempted to prove the *possibility* of freedom. The kind of possibility at stake here is "real" or metaphysical possibility. Mere concepts cannot demonstrate "possibility of any real ground." The fact that the concept of transcendental freedom contains no contradiction is only enough for logical possibility. He frequently issues warnings about how we should not succumb to "the deception of substituting logical possibility of the **concept** (since it does not contradict itself) for the transcendental possibility of **things**" (A244 / B302). He insists that "the possibility of a thing can never be proved merely through the non-contradictoriness of a concept of it, but only by vouching for it with an intuition corresponding to this concept" (B308). Thus, even if he has succeeded in demonstrating that the concept of transcendental freedom contains no contradiction, this would not mean that he has demonstrated its metaphysical possibility.

[114] This is part of the idea behind Kant's famous critique of the ontological argument for God's existence: we cannot infer the existence of a divine being from mere concepts.

The third point that must be taken from Kant's final remark on the antinomy is that freedom "is treated here only as a transcendental idea." In the previous section, it was mentioned that transcendental ideas can employed legitimately as long as they are treated as "regulative" concepts of reason rather than thinking of them as "constitutive." Kant makes that point here as well:

> [T]he transcendental ideas are never of constitutive use, so that the concepts of certain objects would thereby be given, and in case one so understands them, they are merely sophistical (dialectical) concepts. On the contrary, however, they have an excellent and indispensably necessary regulative use, namely that of directing the understanding to a certain goal respecting which the lines of direction of all its rules converge at one point, which, although it is only an idea (*focus imaginarius*) — i.e., a point from which the concepts of the understanding do not really proceed, since it lies entirely outside the bounds of possible experience — nonetheless still serves to obtain for these concepts the greatest unity alongside the greatest extension. (A644 / B672)

As we learned in the discussion of the Paralogisms, the transcendental idea of the soul has value insofar as it provides an orienting point (a *focus imaginarius*) that guides our psychological investigations. We never encounter any experience that vouches for the reality of the soul, but we are justified in positing it as a regulative idea (i.e., as an idea that organizes our psychological questions by giving them a unified subject). Kant thinks there is a natural temptation to treat this regulative idea as a constitutive one and think that a real object is given and that we can have determinate knowledge of it.

But if we avoid this temptation, then we can make legitimate use of the transcendental ideas. When we think of transcendental ideas in terms of their regulative use, we see that their value consists in our capacity to create "subjective principles" or maxims of reason:

> I call all subjective principles that are taken not from the constitution of the object but from the interest of reason in regard to a certain possible perfection of the cognition of this object, **maxims** of reason. Thus there are maxims of speculative reason, which rest solely on reason's speculative interest, even though it may seem as if they were objective principles.
>
> If merely regulative principles are considered as constitutive, then as objective principles they can be in conflict; but if one considers them merely as

> **maxims**, then it is not a true conflict, but it is merely a different interest of reason that causes a divorce between ways of thinking. (A666 / B694).

If we treat the regulative idea of freedom as a constitutive one, then there would be a real conflict between nature and freedom. Rather than treating freedom as a way of *thinking* about noumenal beings in terms of intelligible causes, we would assert freedom as if we could *cognize* it within the empirical world. But it is impossible for us to encounter transcendental freedom in the world of experience.[115]

Freedom, however, stands out from the other transcendental ideas. Our ideas of God and the soul have a regulative value in *scientific investigations*. The idea of God allows us to think of the world as the product of an intelligent author, and this licenses us to think about nature in teleological terms. And Kant thinks there is real value in teleological assumptions, especially in contexts like biology where thinking about bodily organs in terms of their purpose can be illuminating. Nevertheless, Kant warns us that we must treat this idea as a mere regulative principle; we are to think of organs "**as if**" God had given them to an organism for a reason (A688 / B716). The same is true of our idea of the soul. So long as it is taken as a regulative idea orienting our psychological questions, it is legitimate and indispensable.

Freedom stands out from these ideas because it does not have any value whatsoever in the scientific realm. Positing freedom as a cause of events in the world would be a serious stumbling block for scientific inquiry. When thinking about human behavior scientifically, freedom is a

[115] One interesting consequence of this is that we can never be certain that any action was the product of transcendental freedom. Kant admits this in a footnote where he says that "the real morality of actions (their merit and guilt), even that of our own conduct, therefore remains entirely hidden from us. Our imputations can be referred only to the empirical character. How much of it is to be ascribed to mere nature and innocent defects of temperament or to its happy constitution (*merito fortunae*) this no one can discover, and hence no one can judge it with complete justice" (A551 / B579). Although it is possible for intelligible character to have effects on empirical character, we can never be certain that an action came from the former. A good deed (even our own) might simply be the result of a good upbringing; we have no way of knowing its true source.

problem. Nevertheless, like the other transcendental ideas, it does have a legitimate regulative use. The difference is that freedom's proper function is practical.[116] In the case of the malicious lie, freedom entered the picture only as a way of making sense of the moral blame we ascribe to the liar: "this blame is grounded on the law of reason, which regards reason as a cause that, regardless of all the empirical conditions just named, could have and ought to have determined the conduct of the person to be other than it is" (A555 / B583). And in final his discussion of the regulative use of our transcendental ideas, Kant singles out freedom for its practical significance:

> The absolute totality of the series of these conditions in the derivation of their members is an idea which of course can never come about fully in the empirical use of reason, but nevertheless serves as a rule for the way we ought to proceed in regard to them: namely that in the explanation of given appearances (in a regress or ascent), we ought to proceed as if the series were in itself infinite, i.e., proceed *in indefinitum*; but *where reason itself is considered as the determining cause* (*in the case of freedom*)*, hence in the case of practical principles*, we should proceed **as if** we did not have before us an object of sense but one of pure understanding, where the conditions can no longer be posited in the series of appearances, but are posited outside it, and the series of states can be regarded as if it began absolutely (through an intelligible cause); all this proves that the cosmological ideas are nothing but regulative principles and are far from positing, as it were constitutively, an actual totality in such series. (A685 / B713; italics added)

When we are concerned with providing scientific explanations for human actions, they must be thought of as events in an infinitely long chain of causes and effects, which follow necessarily from the laws of nature. But if we consider the actions from a practical perspective, then freedom comes into the picture, and Kant says we should "proceed **as if**" we were not looking at an object of sense but rather at an intelligible cause, which is capable of beginning a causal series

[116] Henry Allison says, "one might expect Kant to claim that the transcendental idea of freedom plays a similar role with respect to pragmatic anthropology. In other words, the claim that we introduce systematic order into our empirical account of human behavior by viewing such behavior 'as if' it were the product of spontaneous, unconditioned acts of free choice. Clearly, however, Kant could allow no such function since it conflicts with his overriding thesis that there is no room even for the thought of such freedom in connection with the empirical character of rational agents. Thus, if the transcendental idea of freedom is to have a regulative function, it must be broadly practical, although not exclusively morally practical, in nature . . . In its regulative, practical function then, what this idea licenses is just the thought of nonempirical conditions in light of which actions may be imputed" (*Kant's Theory of Freedom*, 46).

spontaneously. In the case of Arnold and the shark, if we ask *why it is the case* the Arnold grabs the shark, then our answer should be limited to the causal determinism of the dream machine. Every event in the dream follows those laws. If, however, we ask whether or not it *should be the case* that Arnold grabbed the shark, then we must remove our inquiry from the dream world entirely and deal instead with the question of whether or not Real-Arnold told the neuroscientist to program the machine in that way. It will become clear in the next chapter that, for Kant, questions about moral oughts and rationality cannot find satisfying answers in the empirical world. Thus, thinking about freedom in terms of intelligible causes gives us a domain in which we can think about morality and freedom without (a) violating the laws of nature or (b) "naturalizing" morality by finding it a home in the empirical world.[117]

The fourth and final point to be drawn from Kant's last word on the antinomy is that freedom (properly understood) does not conflict with nature. If we think about transcendental freedom in terms of an intelligible cause that does not interrupt the laws of nature, then there is no conflict. Transcendental idealism has the resources to give libertarian free will a home, and this is why Kant thinks that he is the first philosopher to ever truly reconcile determinism and freedom.

Those are the four conclusions that we must draw from Kant's concluding remark about the antinomy:

> (1) The account about the dual character of the human being does not prove the reality of transcendental freedom.
>
> (2) This account does not prove the metaphysical (i.e., real) possibility of transcendental freedom.
>
> (3) Freedom is a transcendental idea whose only value is the regulative function it serves in the practical domain.

[117] Either one of these would be a serious problem for Kant. The first would jeopardize science by abandoning the causal principle that gives unity to nature. In the next chapter, I explain why (b) would be a problem for morality as it would make moral imperatives heteronomous. For Kant, morality must not be empirically grounded.

> (4) Transcendental freedom (if there is such a thing) does not conflict with the determinism of nature.

Again, we are left with a "problematic" status. Although the concept of freedom contains no contradiction (either with itself or with the concept of nature), we should remain agnostic about its reality and its metaphysical possibility.[118] The resolution of the Third Antinomy shows us only that freedom and nature not logically incompatible. It is crucial that we avoid the mistake of thinking that Kant has provided us with a positive account of how transcendental freedom works. He has not done that, and he *cannot* do that. To do so would be at odds with his many statements of what I call the Incomprehensibility Thesis. Even in the first *Critique*, Kant reminds us that we are not entitled to a positive account of how freedom works: "How such a faculty [*viz*., transcendental freedom] is possible is not so necessary to answer" (A448 / B476). Allen Wood compares Kant's efforts to that of a defense attorney. The burden of proof is such that the prosecution must prove beyond the shadow of a reasonable doubt that the defendant is guilty. Kant's only aim is to show that freedom might be innocent of the charges leveled against it by its accusers. He makes the minimally ambitious claim that it cannot be proven that freedom is logically impossible, and it cannot be demonstrated that it conflicts with nature. The positive story he conjures up is like an alibi, and the defense only needs to show that the alibi casts doubt on the prosecution's case:

> The attorney need not show that his theory is the correct one, only that it has enough plausibility to introduce a reasonable doubt concerning the prosecution's theory. Likewise, in defending our freedom, Kant concocts a metaphysical theory which, if true, saves our practical freedom despite the fact that our actions are determined by natural causes. Kant does not need to show that this metaphysical theory is correct; indeed he frankly admits that he cannot. But neither, he claims, can the

[118] We should remain agnostic *at this point*. Theoretical reason offers us no justification for a belief in transcendental freedom. But, as I explain in the next chapter, practical reason does give us a compelling reason to believe that we are free.

> opponent of freedom refute the theory. Kant rests his defense of freedom on this claim.[119]

Viewed in this light, Kant's theory of transcendental freedom might get a pass. He does not have to demonstrate that his view is terribly plausible; by his own admission, he can't even prove that it's *possible*. All he has to prove is that it cannot be shown to be impossible, and that it does not necessarily conflict with nature. This is the fifth and final point to be concluded from the Third Antinomy; the account he gives does not necessarily reflect a positive view of how freedom actually works. It is, at most, a story that cannot be proven to be false. As long as one is willing to entertain the possibility of transcendental idealism, transcendental freedom cannot be dismissed.

2.5 Conclusion

Freedom comes up in the first *Critique* because reason refuses to confine itself within the boundaries of possible experience: "Reason is driven by a propensity of its nature to go beyond its use in experience, to venture to the outermost bounds of all cognition by means of mere ideas in a pure use, and to find peace only in the completion of its circle in a self-subsisting systematic whole" (A797 / B825). In its pursuit of an all-encompassing unity and systematicity, reason begins to think that an unconditioned condition must be given for the conditioned things that are presented to it. But this gives rise to unwarranted metaphysical speculation.

We might be left wondering why Kant thinks we are still prone to these metaphysical errors after reading the *Critique.* After all, his aim in the Dialectic is to expose the root of these errors and to discipline reason's metaphysical pretensions. We are supposed to realize that these errors involve the fallacies of equivocation, assumptions of transcendental realism, or the conflation of regulative and constitutive concepts. Shouldn't we have learned how to always avoid these

[119] Wood, "Kant's Compatibilism," 248.

pitfalls? Kant's answer is negative. Although we can expose the errors behind transcendental illusion, the illusion itself does not cease: "The transcendental dialectic will therefore content itself with uncovering the illusion in transcendental judgments, while at the same time protecting us from being deceived by it; *but it can never bring it about that transcendental illusion (like logical illusion) should ever disappear and cease to be an illusion* . . . so even after we have exposed the mirage it will not cease to lead our reason on with false hopes, continually propelling it into momentary aberrations that always need to be removed" (A298 / B354; emphasis added). In this way, transcendental illusion is like an optical illusion in which there are two lines of equal length but one seems longer than the other. Even after measuring the lengths, it will always *look* like one line is shorter. Transcendental idealism gives us a ruler, and it exposes the source of the illusion. But it cannot eradicate the deceptive appearance. Kant believes we are irresistibly tempted to regard certain appearances in the world as products of freedom even though we have no way of comprehending how this is actually possible.[120]

When it comes to the existence of God, the properties of the soul, and the nature of the world as a totality, we cannot help but be led astray. Kant thinks we will inevitably want to regard these ideas as given, and we are inclined to conflate their proper regulative use with an improper constitutive use. We find ourselves contemplating these transcendental ideas only after transcending the "bounds of sense," leaving behind the terra firma of possible experience. Kant asks us to do our best to remember this fact even when we are lost at sea on the "broad and stormy ocean" of illusion. In the case of transcendental freedom, we must not forget that theoretical reason offers no justification for the belief that we are free or that freedom is really possible. For reasons that I explore more fully in chapter 4, speculative reason can offer no positive account of how

[120] As I discuss at length in the next two chapters, our practical view of the world requires this belief; it is theoretical reason that has trouble with it.

freedom actually works. What it can and must do is make enough room for the possibility of freedom that the idea cannot be definitively ruled out. An essential part of that project is showing that freedom does not conflict with determinism. This has two important implications for the unity of Kant's critical philosophy. First, it means that the antinomy does not pose any real threat to theoretical reason. Second, and just as important, this lays the groundwork for his practical philosophy. As I explain in the next chapter, transcendental freedom is a necessary condition of Kant's moral philosophy. If the first *Critique* had demonstrated that freedom is impossible, then reason's practical vocation — the proper exercise of which constitutes the human being's ultimate end — would be hopeless from the start.

Chapter 3
Freedom in the Practical Philosophy

3.1 Introduction

In the last chapter, I explained Kant's view that theoretical reason can offer no justification for the existence or even the metaphysical possibility of transcendental freedom. But within the domain of practical reason, Kant settles on what he takes to be definitive reasons for asserting that human beings do in fact have transcendental freedom. As I explained in chapter 1, autonomy or moral freedom (i.e., strong practical freedom) requires transcendental freedom. We are immediately aware of the fact that we are bound by the moral law, and, since the moral law requires freedom, Kant thinks can know that we are free. As he puts it, "[T]he moral law is the *ratio cognoscendi* of freedom" (*KpV* 4).

It is natural to wonder what exactly it is about the moral law (or our consciousness of it) that allows us to claim that we possess transcendental freedom. Unfortunately, the details of that connection are rather controversial, and many commentators are unsatisfied with the argument. Derk Pereboom says that the practical reasons Kant offers in favor of freedom are "subject to serious challenge," Paul Guyer dismisses the fact of reason argument as "foot-stamping," calling it "one of the most spectacular train wrecks," and Allen Wood calls it a "moralistic bluster."[121] They are quick to point out that the argument from morality to freedom is "widely considered an abysmal failure."[122] Since so much of Kant's philosophy rides on this conclusion, it would be a great disappointment if his argument for it turns out to be nothing more than thinly veiled

[121] Pereboom, "Transcendental Freedom," 537. Guyer, "Naturalistic and Transcendental Moments", 462. Wood, *Kantian Ethics*, 135.

[122] Owen Ware, "Rethinking Kant's Fact of Reason," 1.

dogmatism. My aim in this chapter is to offer a novel, charitable reconstruction of Kant's view.[123] But we should not find the scholarly dissatisfaction with the argument terribly surprising; when it is presented in its simplest form, it certainly seems dubious. It goes as follows:

(1) We are bound by the moral law.
(2) We are bound by the moral law only if we possess transcendental freedom.
Therefore,
(3) We possess transcendental freedom.

Of course, both premises are in need of further explanation and justification, and I explore each premise in detail below. First, in section 3.2, I explain what the moral law is, what it means to be bound by it, and what justifies our belief in its bindingness. Kant's justification for the claim that we are bound by the moral law is his "fact of reason" argument, which I have broken down into in three subsections. I explain what I take to be the referent of the fact in 3.2.1. Since the argument makes use of an analogy with theoretical reason, I provide some necessary background information about theoretical reason in 3.2.2. I complete the analogy in 3.2.3 by showing how the claim that we are bound by the moral law lives up to the criteria for a priority that Kant uses in the theoretical domain. I address the second premise (that morality requires transcendental freedom) in section 3.3. I present an original reconstruction of Kant's argument, and I underscore the importance of moral feeling/respect, which I believe many commentators have not fully appreciated. In broad strokes, this is how I reconstruct the argument: When we engage in introspection, we are immediately conscious of the existence of categorical moral prescriptions (oughts) that come to us a priori from a non-empirical source (*viz*., pure reason). In addition to these categorical oughts, we are also aware of the feeling of respect, which Kant takes to be an empirical, sensible effect of

[123] To the best of my knowledge, no one reconstructs the argument in the way that I present it here. I explain below why I think other commentators are mistaken, and I believe that some of the scholarly dissatisfaction with Kant's view can be allayed if we do not rely on the unsatisfactory versions of the argument that can be found in the literature.

pure reason's capacity to determine the will by means of those oughts. Kant believes that if he is successful in demonstrating those claims, which amount to the conclusion that "pure reason can be practical," then he has shown that we have transcendental freedom.

3.2 Bound by the Moral Law

For Kant, the moral law is expressed by the categorical imperative, and although he says that there is "only a single categorical imperative," he gives several formulations of it (*G* 4:421). The first version of the categorical imperative is the formula of universal law (FUL): "act only in accordance with that maxim through which you can at the same time will that it become a universal law" (*G* 4:421). This prescribes the following decision procedure.[124] The first step is to articulate the maxim[125] you intend to act on. The second step is to imagine a world in which every rational agent acted on that maxim. Finally, you must ask yourself whether or not such a world is conceivable and if you could "rationally *will* to act on your maxim in such a world."[126] If such a world is inconceivable, then you have a "perfect duty" (i.e., one "that admits no exception in favor of inclination" [*G* 4:421]) to refrain from acting on that maxim. For instance, there is a perfect duty to refrain from making false promises since a world in which everyone makes false promises would be inconceivable as there would be no promises in such a world. And if the world is

[124] Many commentators follow O'Neill and Rawls in reading the FUL as a decision procedure. I also adopt this practice since it seems like the most natural interpretation. See O'Neill, Onora *Acting on Principle* and *Constructions of Reason.* Cf. Rawls, *Lectures on the History of Moral Philosophy*, and "Themes in Kant's Moral Philosophy."

[125] Kant defines a maxim as the "subjective principle of volition" (G 4:402), but this definition requires some elaboration. Fortunately, Kant offers a handful of examples from which we can derive the basic ingredients of maxims. Here are two of the examples: "[F]rom self-love I make it my principle to shorten my life when its longer duration threatens more troubles than it promises agreeableness." "[W]hen I believe myself to be in need of money I shall borrow money and promise to repay it, even though I know that this will never happen" (G 4:422). In every example, there is some type of act A, some circumstance C, and some end E. As Robert Johnson puts it, the form of a maxim is: "I will *A* in *C* in order to realize or produce *E*" (Johnson, "Kant's Moral Philosophy."

[126] Johnson, "Kant's Moral Philosophy."

conceivable but you cannot rationally will to live in such a world, then you have an "imperfect duty" (i.e., one which admits of some exceptions due to differing inclinations) to refrain from acting on that maxim.[127] If the world is conceivable and you can will to live in that world, then it is morally permissible to act on that maxim.

The second formula is the formula of humanity as an end in itself (FH): "So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means" (*G* 4:429). This formula is somewhat more straightforward than the first. It simply demands that we never treat a human being (ourselves included) like a mere object, for that would be at odds with the unassailable dignity that we should attribute to all rational beings. Of course, this does not mean that no one can ever be used a means. Rather, we must never treat anyone as a *mere* means; that is, we must also always respect them as an end. When considering the case of the false promise to the banker, it is clear that this reduces the banker to a mere means of acquiring money, and thus the false promise is impermissible; whereas, if the promise is kept and the loan is repaid, then although the banker was a means of acquiring money, she was not *merely* a means.

The third formula is remarkably similar to the first. It is usually referred to as the formula of autonomy (FA), and it is "the idea of the will of every rational being as *a will giving universal law*." (*G* 4:432). The principal difference between FA and FUL is that the former depicts us as law*givers* and the latter as law *followers*.[128] The final formula, the kingdom ends (FKE), synthesizes these two components by asking us to think of ourselves both as a legislator and as a

[127] Kant gives an example about idleness and cultivating talents. Although a world in which no one cultivates talents is conceivable, one cannot rationally want to live in such a world.

[128] See Johnson, "Kant's Moral Philosophy."

subject in a hypothetical kingdom of ends. That is, we should "act in accordance with the maxims of a member giving universal laws for a merely possible kingdom of ends" (4:439).[129]

Kant famously says that "the above three ways of representing the principle of morality [FUL, FH, and FKE] are at bottom only so many formulae of the very same law, and any one of them of itself unites the other two in it" (*G* 4:436). However, in spite of the claim that each formula "unites the other two in it," the equivalence of the formulas should not be thought of as an analytic connection. It would be quite a stretch to think of FUL as analytically containing FH or vice versa. The most plausible interpretation (and the one which I favor here) is that the different formulas will never generate conflicting duties. Any maxim that passes one test will pass the others, and any maxim that fails one test should fail the others.

That explains what the moral law is, but it does not explain what it means for us to be "bound" by it. It is here that certain details about the constitution of human beings become relevant. Although the moral law is the same for all rational beings (human and nonhuman alike)[130], its relation to the will varies depending on the nature of the agent.

> 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. However, if reason solely by itself does not

[129] As Johnson puts it, "The intuitive idea behind this formulation is that our fundamental moral obligation is to act only on principles which could earn acceptance by a community of fully rational agents each of whom have an equal share in legislating these principles for their community."

[130] Kant frequently makes this point about rational beings. It is not clear what (if any) nonhuman rational beings he has in mind (perhaps angels or extraterrestrials). The point, however, is an important one. Much of his argument rides on the claim that the moral law springs from rationality itself, not from the particular constitution of human beings. So the moral law must be the same for all rational beings; it cannot depend on the particularities of human beings.

It is interesting to note (though not central to the argument) that Kant does make a couple references to extraterrestrials in his *Anthropology*: "The highest concept of species may be that of a terrestrial rational being [*eines irdischen vernünftigen*], but we will not be able to describe its characteristics because we do not know of a nonterrestrial rational being [*nichtirdischen Wesen*] which would enable us to refer to its properties and consequently classify that nonterrestrial being as rational. It seems, therefore, that the problem of giving an account of the character of the human species is quite insoluble [*sie schlechterdings unauflöslich*], because the problem could only be solved by comparing two species of rational beings on the basis of experience, but experience has not offered us a comparison between two species of rational beings" (*A* 8:215).

> 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 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*: that is to say, the relation of objective laws to a will that is not thoroughly good is represented as the determination of the will of a rational being through grounds of reason, indeed, but grounds to which this will is not by its nature necessarily obedient. The representation of an objective principle, insofar as it is necessitating for a will, is called a command (of reason), and the formula of the command is called 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 constitution is not necessarily determined by it (a necessitation). (*G* 4:413–14)

Thus, for human beings, the moral law binds us in the sense that we always *ought* to act in accordance with the categorical imperative.[131] This relation would not obtain, however, for a perfectly rational being (like an angel who is not tempted by inclinations). For as Kant argues, "[N]o 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" (*G* 4:414). The angel's will would necessarily conform to the moral law, but humans are tempted by sensuous desires, and so the human will does not necessarily act in accordance with the moral law. To say that human beings are bound by the moral law is to say that they *ought* to act in accordance with it.

Of course, replacing the language of 'the moral law binding our will' with talk of 'oughts' only pushes the question back a step. There must be some further explanation of what it means to say that a human being 'ought' to act in accordance with the moral law. The best way to demystify Kant's concept of 'the moral ought' is to think of it as generating an unconditional, overriding reason for acting a certain way (*viz*., acting only on maxims which past the tests described above). The unconditional nature of these reasons (i.e., the fact that they are reasons for *all rational agents*

[131] And, of course, if we are to achieve moral worth, then we must not only act in accordance with the moral law but we must also act from the motive of duty. See *Groundwork* 4:396–98.

at *all* times) is what distinguishes them from reasons generated by prudential oughts, which are reasons for action only insofar as some end is desired. But it is also crucial that the reasons be overriding. Again, in Kant's view, morality takes the form of 'oughts' only when moral reasons have to compete with non-moral ones, and moral reasons *should* override others because, from the perspective of rationality, they are authoritative.[132] That is why the 'ought' is out of place when discussing a holy will. There are no competing inclinations, and thus it does not make sense to say that a perfectly rational being ought to act in accordance with the moral law because such a being will always act morally. So, at bottom, the claim that "we are bound by the moral law" amounts to the claim that we always have reasons to act in accordance with the categorical imperative, and to act contrary to the dictates of these reasons is to forfeit our autonomy and act irrationally, since these reasons *should* override the competing, heteronomous incentives of self-love.

3.2.1 The Fact of Reason

But what is the justification for this? How can we be so sure that morality is binding for us in this way? Unfortunately, Kant spills very little ink to back up this claim,[133] and what little he does say is obscure and has been the source of much scholarly controversy. In the *Critique of Practical Reason*, he boldly asserts that our consciousness of the moral law is a "fact of reason."

> It is therefore the *moral law*, of which we become immediately conscious (as soon as we draw up maxims of the will for ourselves), that *first* offers itself to us and, inasmuch as reason presents it as a determining ground not to be outweighed by any sensible conditions and indeed quite independent of them, leads directly to the

[132] One way to understand the sense in which moral reasons are overriding is by an analogy with rational egoism. An egoist might recognize that the pleasure of smoking a cigarette constitutes a reason for smoking, but her health and future well-being give her a reason not to smoke. Insofar as she values her longevity more than the fleeting pleasure of smoking, her health-driven reasons are overriding. That is, from the all-things-considered perspective of her rationality, she should not smoke. Kant's moral reasons work in a similar way. Although inclinations of self-love might look like reasons for actions, moral reasons must take precedence if the agent is to act rationally.

[133] Much of the controversy arises from the problem that this central premise of the second *Critique* (the fact of reason) is given a scandalously brief explanation.

> concept of freedom. But how is consciousness of that moral law possible? We can become aware of pure practical laws just as we are aware of pure theoretical principles, by attending to the necessity [*Notwendigkeit*] with which reason prescribes them to us and to the setting aside of all empirical conditions to which reason directs us.[134]
>
> 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, although it would be analytic if the freedom of the will were presupposed; but for this, as a positive concept, an intellectual intuition would be required, which certainly cannot be assumed here. However, in order to avoid misinterpretation in regarding this law as *given,* it must be noted carefully that it is not an empirical fact but the sole fact of pure reason which, by it, announces itself as originally lawgiving (*sic volo*, *sic jubeo*).[135]

At first glance, it might look as if Kant is asking the reader to accept that our consciousness of the moral law is a brute fact that admits of no further explanation. Alternatively, he might be resorting to an alleged fact about human psychology (e.g., people feel guilty when they do something wrong, we know right from wrong, etc.) It is these lines of thinking that have troubled so many readers.[136] It seems to them that the second *Critique* begs its most important question. As Owen Ware puts it, "[M]any commentators in Kant's day and our own have found the second *Critique* disappointing because the book's argument begins where they think it should end: with our consciousness of the moral law."[137] But there are subtle hints from the above excerpts that suggest a more charitable interpretation.

The most noteworthy clue is the analogy between the fact of reason and pure theoretical principles. Kant says that we become of aware of the moral law in just the same way that we become aware of theoretical principles; we accomplish this task "by attending to the necessity

---

[134] *KpV* 5:29–30
[135] *KpV* 5:31
[136] See, for example, Guyer, "Naturalistic and Transcendental Moments," 462. Cf. Wood, *Kantian Ethics*, 135.
[137] Ware, "Rethinking Kant's Fact of Reason," 16.

[*Notwendigkeit*] with which reason prescribes them to us and to the setting aside of all empirical conditions to which reason directs us" (*KpV* 5:30). These two criteria of the a priori (*viz*., necessity and purity[138]) pervade Kant's theoretical philosophy. His theoretical project hinges on providing a satisfactory explanation of how synthetic a priori judgments are possible. If it turns out that a synthetic judgment (e.g., the shortest distance between two points is a straight line) either lacks or necessity or has its basis/ground in experience, then the judgment could not be a priori. So in order to understand the fact of reason and the comparison with theoretical principles, we must first examine some of the a priority arguments from the *Critique of Pure Reason*.

But before comparing the fact of reason with theoretical principles, it is important to be clear about what exactly the 'fact' is. Kant's presentation of it varies from place to place, and his choice of wording has led to two significant disputes. First, there are differing opinions about the referent of the fact. It is not immediately obvious whether Kant is referring to our consciousness of the moral law or to the moral law itself. There are eight references[139] to the fact of reason in the second *Critique*. The first time he uses the phrase '*Faktum der Vernunft*,' he is referring to "our consciousness of the moral law" (5:31), but, immediately after the first mention of it (indeed, on the very same page), he says that the fact refers to the moral law itself — not our consciousness of it. The third instance is perhaps the most surprising, as he says the fact is the "autonomy in the principle of morality by which reason determines the will to deeds" (5:42). And the fourth use refers once more to the law: "[T]he moral law is given, as it were, as a fact of pure reason of which we are a priori conscious and which is apodictically certain" (5:47). So which is it? Does the fact

[138] Throughout this chapter, 'purity' refers to non-empirical origin/justification.

[139] For my purposes here, there are four references to the fact of reason in which Kant offers some suggestion as to its referent. There are other places where he mentions the fact in passing without saying almost anything about what the fact might be. For instance, he says in the preface, "[N]ow practical reason of itself, without any collusion with speculative reason, furnishes reality to a supersensible object of the category of causality, namely to *freedom* (although, as a practical concept, only for practical use), and hence establishes **by means of a fact** what could there only be *thought*" (*KpV* 5:6; boldface added).

refer to our consciousness of the moral law, the existence of the law, or the autonomy of the principle of morality?

The second question is how to interpret '*Faktum*.' This is usually translated into English as 'fact,' but it is linguistically derived from the Latin 'facere,' which means 'to do' or 'to make.' And thus it is reasonable to think that '*Faktum*' is not simply a 'fact' in the sense of a state of affairs but rather a 'deed' — something that has been done. Interpreting '*Faktum*' as 'deed' is especially plausible given that Kant could have used the word '*Tatsache*,' which also means 'fact,' but is distinct from '*Tat*,' which means 'deed.' Furthermore, Kant specifically draws on this legal meaning (i.e. '*factum*' as 'deed') in the Transcendental Deduction of the first Critique. He begins by comparing his deduction of the categories to a legal matter in which the jurists must distinguish between "questions about what is lawful (*quid juris*) and that which concerns the fact (*quid facti*)" (A84/B116). In a criminal trial the jurors must determine both what was done (*quid facti*) and whether or not it was lawful (*quid juris*).

These are not trifling semantic disputes; the stakes are actually quite high. The soundness of the second *Critique*'s central argument stands or falls depending on which interpretation we choose. If the fact amounts to nothing more than the claim that human beings are usually conscious of a distinction between right and wrong (and they are aware that what morality asks of them is sometimes at odds with their self-interest), then the fact might be plausible, but it surely could not bear the argumentative weight that Kant places on it. We could not infer from such a fact that we have transcendental freedom. Another potential problem is that the legal sense of '*Faktum*' (as a deed) seems to imply that it requires further proof, even though Kant says that the moral law "cannot be proved by any deduction" or "confirmed by experience" (*KpV* 5:47). As Owen Ware points out, "Quite distinct from the meaning they would later acquire, facts were not objects of

reasonable belief, but items of evidence to be determined by a jury . . . [I]f we interpret Kant's *Faktum* solely in terms of an analogy to eighteenth-century criminal law, we are forced to demote the status of moral consciousness to something allegedly the case, so as to make Kant's *Faktum* fit the idea of legal evidence bearing 'defeasible, non-demonstrative force.' To do so, however, is completely at odds with the text and spirit of the second *Critique*, where Kant repeatedly affirms that the moral law is "apodictically certain" and "firmly established of itself" (*KpV*, AA 5:47).[140]

Surprising as it might sound, the solution here is understand how all of these seemingly divergent interpretations can be united. When it comes to the referent of the fact, the best choice is to understand it as the fact that we know we are bound by the moral law. It was explained above in section 3.2 that this amounts to the claim that we always have unconditional, overriding reasons to act in accordance with the moral law. If we take this as the referent, we can see how Kant's three descriptions above are not at odds with one another. They do not need to compete at all. Once again, the three referents in the text are (1) the fact that we are conscious of the moral law, (2) the fact that there is a law that comes from pure reason alone, and (3) the autonomy of the principle of morality through which it determines the will. If we understand the fact of reason in terms of our awareness that reason generates moral oughts, we can see how (1) must be true. Our awareness of these oughts is nothing other than consciousness of the moral law. There is, however, an obstacle in seeing how are supposed to move from (1) our consciousness to (2) the purity of the law's origin and (3) the autonomy of the principle of morality.

Indeed, that move is one that has troubled scholars for some time. In his influential commentary on the fact of reason,[141] Lewis White Beck puts the problem quite forcefully: "Kant

[140] Ware, "Rethinking Kant's Fact of Reason," 16.

[141] Allison says that Beck's "analysis is the mandatory starting point for any serious discussion of the fact of reason" (*Kant's Theory of Freedom*, 232).

cannot make a transition from the undisputed fact (that we are conscious of a moral law) to the disputed fact (that there is a law that can only come from pure practical reason). Yet it is the latter whose factuality is to be shown if the moral law is to be justified."[142] But this difficulty is dissolved when we take the fact to be the bindingness of the moral law and the consciousness to be merely a necessary condition of that fact.[143] The task is no longer to move from our consciousness of a distinction between right and wrong to the claim that pure reason is the author of the law. Instead, we must move from the fact that we have unconditional reasons to act morally to the claim that these reasons must have come from pure reason and not from an empirical source. And the latter move is far more plausible than the former. Of course, Kant still needs to justify the claim that these reasons (i.e., moral oughts) must have come from pure reason and could not have an empirical ground. If he can show that, then (2) and (3) are also comprised in the fact of reason. That argument is presented below in section 3.2.3.

But before moving on, we should also address the question of whether to read '*Faktum*' as a 'deed' or a 'state of affairs,' I believe that we can and should retain both meanings. On the one hand, Kant is certainly talking about a deed [*Tat*] that reason accomplishes, namely, it the activity of generating unconditional reasons (moral oughts) for action. On the other hand, our awareness of this activity is a state of affairs [*Tatsache*] which needs no further evidence. Many commentators share this interpretation,[144] but Marcus Willaschek's analysis is especially illuminating:

---

[142] Beck, *A Commentary on Kant's* Critique of Practical Reason, 168.

[143] This does raise another problem, however. Although it might remove the difficulty that Beck points out (moving from an uncontroversial fact to a controversial one), it might look like Kant is assuming the legitimacy of the fact of reason without argument and thus begging the question of the second *Critique.* I hope to assuage those concerns in my reconstruction of the argument below.

[144] But, as always, there is no consensus on the matter. Owen Ware argues that Fact of Reason should be seen as a kind of '*Tatsache.*' And he points out that he is not alone in this regard. Cf. Kleingeld, "Moral Consciousness." But Ware argues that his interpretation is compatible with reading 'Faktum' as 'act' or 'deed.' He writes, "I believe we can reconcile these readings if we maintain that moral consciousness only arises through an original act of reason's self-determination. This would still make moral *consciousness* the primary referent of Kant's *Faktum*, but it would

> This ambiguity makes it as evident as it can be that the "fact of reason" refers to a deed [*Tat*] and also a fact [*Tatsache*], however, not a fact as something merely 'given' (*datum*), but a 'deed-thing' [*Tat-Sache*] as the result of a deed [*Tat*] (*factum*). Even when the term seems to imply something like 'fact,' it is not necessary to understand it in the sense of a brute fact, in which reason simply finds this '*Factum*'; rather, the fact, the '*Factum*,' is always a product of reason.[145]

The benefit of understanding '*Faktum*' as an activity that reason performs is that its source would no longer be a mystery. We do not discover the fact by means of some kind of intellectual intuition (which would violate Kantian strictures). The *Faktum der Vernunft* refers simply to the fact that reason generates moral oughts (the *Tat*), and we are immediately aware of this state of affairs (the *Tatsache*).[146] The danger associated with thinking of '*Faktum*' strictly as a deed is that, in the legal sense, a *factum* would require further evidence, but the *Faktum der Vernunft*, like Descartes's *cogito* is foundational and cannot be deduced from or proven by anything else.[147]

On the reading I am suggesting, the "fact of reason" does not beg the question of the second *Critique*. Even if the reader grants Kant the above claims about the fact of reason, the principal task (showing that pure reason can be practical)[148] is not thereby finished.[149] What has been granted

---

preserve the important insight by Willaschek, Engstrom, and Franks that the moral law is one we actively give to ourselves." Ware, "Rethinking Kant's Fact of Reason," 2.

[145] Willaschek, *Praktische Vernunft: Handlungstheorie und Moralbegründung bei Kant*, 181; translation mine.

[146] This also allows us to see how Kant's remark in the preface anticipates the fact of reason and the role that it will play in establishing the reality of freedom. He says that "if as pure reason it is really practical, it proves its reality and that of its concepts by what it does [*durch die Tat*]" (5:3).

[147] Nevertheless, like Descartes's *cogito*, the foundational nature of the fact of reason does not mean that it cannot be explained further or illustrated. Descartes uses the evil deceiver to show how it is impossible to doubt our existence, but that does *not* make the *cogito* the conclusion of a logical deduction, because if that were so, then it could be doubted insofar as the tools of deduction could be doubted. So it is with the fact of reason. Although it is foundational and cannot be deduced from anything, that does not mean that it cannot be illustrated. It will become clear in what follows that Kant uses thought experiments to illustrate (though not deduce) the fact of reason.

[148] This is how Kant puts it throughout the second *Critique*. If I were to take Kantian jargon out of it, this amounts to the claim that that we are capable of acting for the sake of something other than self-interest, namely for the sake of morality. Since Kant takes self-interest to be exhaustive of our empirically grounded interests, if we are capable of acting against our self-interest for the sake of morality (which comes from pure reason), then pure reason must be capable of determining the will.

[149] In this way, I believe Kant can avoid the objection that the fact of reason begs the question of the second *Critique*. The fact of reason does not show that pure reason can be practical. When Kant says that "pure reason can be practical" he means that "pure reason of itself [i.e. independently of inclination] alone suffices to determine the will"

so far is that reason generates unconditional reasons for action and that we are conscious of this activity. It remains to be seen whether or not Kant can justify the claim that these reasons have the force of necessity and that they spring from a pure (i.e., non-empirical) source.[150] It also must be demonstrated that these pure moral oughts can determine the will without the assistance of anything empirical. Those are the questions that will be addressed in the next two sections. Once again, the key to this puzzle lies in seeing how we "can become aware of pure practical laws just as we are aware of pure theoretical principles" (*KpV* 5:30).[151]

3.2.2 The Necessity and Purity of Theoretical Principles

In the Transcendental Aesthetic, Kant maintains that space attaches to "the form of intuition alone" and is thus constitutive of our representations of objects of outer appearance (A23/ B38). As he argues for this conclusion, he claims that (1) "Space is not an empirical concept that has been drawn from outer experiences" and (2) "Space is a necessary representation, a priori, which is the ground of all outer intuitions" (A23–24 / B38).[152]

(5:15). The fact of reason plays a crucial role in getting Kant to that conclusion, but he is not simply assuming the truth of his conclusion as a premise.

[150] It is conceivable that something could be universal but not have an a priori foundation. For instance, humans might universally share some trait (e.g., aversive reactions to certain kinds of painful stimuli), but that universality would be grounded in a shared empirical constitution rather than pure reason. So the fact that moral oughts are universal is not yet enough to demonstrate that they have the force of necessity generated by pure reason.

[151] In one of his notes Kant affirms that this is indeed his strategy in the second *Critique*: "The distinction between empirically-conditioned and pure, yet still practical reason is foundational for the critique of practical reason, which asks if there is such a thing as the latter. Its possibility cannot be comprehended *a priori*, because it concerns the relation of a real ground to its consequent. Something must therefore be given, which can stem only from it; and its possibility can be inferred from this reality. Moral laws are of this nature, and these must be proven in the manner in which we prove that the representations of space and time are *a priori*, with the difference being that the latter are intuitions and the former mere concepts of reason" (R 7201, 19:276–76).

[152] He also argues (3) "The apodictic certainty of all geometrical principles and the possibility of their a priori construction are grounded in this a priori necessity;" (4) "Space is not a discursive or, as is said, general concept of relations of things in general, but a pure intuition'" and (5) "Space is represented as a given infinite magnitude," but these are less germane to the analogy at issue here since the main thrust of the analogy is that the moral law is necessary and cannot come from experience.

In the service of (1), Kant asks the reader to try to imagine how the representation of space could have been learned from experience. The suggestion he offers is to consider two objects which are in different places. Although an empiricist would like to demonstrate that our idea of space comes from experiences like this one, Kant argues that space is a precondition of having such an experience and must therefore be a priori:

> For in order for certain sensations to be related to something outside me (i.e., to something in another place in space from that in which I find myself), thus in order for me to represent them as outside one another, thus not merely as different but as in different places, the representation of space must already be their ground. Thus the representation of space cannot be obtained from the relations of outer appearance through experience, but this outer experience is itself first possible only through this representation. (*KrV* A23/ B38)

This kind of argument runs all throughout the first *Critique*. The idea behind it is that we could not have learned X from experience because X is a condition of the possibility of experience.[153] He takes the point even further by showing that our representation of space not only lacks empirical origins but also that it is *necessary*.

In order to justify (2), the claim of necessity, Kant once again invokes a kind of thought experiment. He says, "One can never represent that there is no space, although one can very well

[153] It is likely that an empiricist would push back against this line of thinking and argue that even if something is a condition of experience, it still must have been learned by means of an experience. For no knowledge of a condition of experience could temporally precede the experience itself. But Kant would readily concede this sort of objection. He says at the outset of the first *Critique* that "There is no doubt whatever that all our cognition begins with experience; for how else should the cognitive faculty be awakened into exercise if not through objects that stimulate our senses . . . **As far as time is concerned**, then, no cognition in us precedes experience, and with experience every cognition begins.

But although all our cognition commences **with** experience, yet it does not on that account all arise **from** experience. For it could well be that even our experiential cognition is a composite of that which we receive through impressions and that which our own cognitive faculty (merely prompted by sensible impressions) provides out of itself, which addition we cannot distinguish from that fundamental material until long practice has made us attentive to it and skilled in separating it out" (B1–2).

And thus Kant would argue that we might learn about space after having experiences of things in space, but that does not mean that our knowledge of space comes **from** those experiences. Kant argues that certain components of experience (space, time, causality, etc.) are, as it were, brought to the experience by the subject. This is why being a precondition of experience suffices to make something a priori. It must be logically prior to the experience if the subject supplies it as a precondition of the experience.

think that there are no objects to be encountered in it. It is therefore to be regarded as the condition of the possibility of appearances, not as a determination dependent on them, and is an a priori representation that necessarily grounds outer appearances" (A24/ B38–39). It is tempting to read this as making a claim about psychological necessity. According to this reading, we as humans can conceive of space with no objects in it, but we cannot conceive of objects without space. The glaring problem with this reading is that it makes the argument a non-sequitur. As Allison puts it, "if it is interpreted as asserting the impossibility, as a matter of psychological fact, of representing to oneself the absence of space, it becomes difficult to see how it could support, much less entail, Kant's conclusion that 'space is an *a priori* representation that *necessarily grounds* outer appearances.'"[154]

We must find an alternative lest we find Kant guilty of making a patently fallacious argument. It has also been suggested that Kant is asserting a relationship of logical necessity between objects and space. "On this interpretation, Kant is making a claim about the inconceivability of the nonexistence of space that is analogous to Spinoza's claim about the inconceivability of the nonexistence of substance."[155] Allison points out that this reading is at odds with much of what Kant says elsewhere. For Kant never makes the claim that space is logically necessary, and he frequently insists that we must allow room for a being (perhaps a divine one) that would be capable of a kind of non-sensible, intellectual intuition. This kind of intuition would not require the forms of sensibility (space and time). So space cannot be logically necessary.

The third and most promising interpretation is that the necessity results from the way in which space is constitutive of outer appearances. Kant's point is not that space is psychologically necessary (i.e., that we cannot help but think of objects as existing in space); nor is he saying that

[154] Allison, *Transcendental Idealism*, 105.
[155] *Ibid*.

space logically necessary. Rather, "space is necessary for the representation of outer appearances."[156] It is not that that we cannot think of objects without space. After all, Kant's system requires that we able to think (though not have intuition of) things in themselves, and since things in themselves are not in space and time, we must surely be able to think of things without representing them as being in space. The necessity comes from how we *represent* outer appearances. Once we remove space from a representation, we remove all of the objects of outer appearance. Kant's idealism requires that objects of appearance depend on thinking subjects' representations for their existence. As he puts it in the first edition, "if I were to take away the thinking subject, the whole corporeal world would have to disappear, as this is nothing but the appearance in the sensibility of our subject and one mode of its representations" (A383).

The representation of space is necessary a priori, and our intuition of it is pure (i.e., not empirical). This is important to Kant because, synthetic a priori judgments about geometry (e.g., the sum of the interior angles of a triangle is equal to two right angles) obviously cannot come from an empirical, contingent source. If we were to ground geometrical judgments in experience, then they could never sustain claims of necessity. [157] We could only say that experience has *thus far* shown us that this judgment holds. It would be contingent and a posteriori. But such judgments are "apodictically certain," and they must have an a priori ground.

---

[156] *Ibid.* Allison calls this necessity 'epistemic,' but that label is perhaps wed to his particular reading of transcendental idealism as a distinction between two different ways of thinking about objects. This kind of necessity is still applicable within a metaphysical interpretation of transcendental idealism, and so I refrain from calling it 'epistemic.'

[157] Kant does not think that our pure intuition of space alone is enough to explain the objective validity of geometry. He must also argue, as he does in the Axioms of Intuition, that the principles of mathematics (which concern pure space and time) apply to objects of appearance that come to us through sensible intuition. He argues there that we represent appearances, "through the same synthesis as that through which space and time in general are determined" (A161 / B202).

There are many other places where Kant makes similar arguments about attributing necessity to some judgment and showing that the judgment cannot have an empirical ground. The concept of substance passes the same test. When we reflect on our judgments about objects, we should be able to strip away everything learned from experience (e.g., "the color, the hardness or softness, the weight, even the impenetrability" [B6]) and find that we cannot remove the concept of substance.[158] As Kant says in the B Introduction "[I]f you remove from your empirical concept of every object, whether corporeal or incorporeal, all those properties of which experience teaches you, you could still not take from it that by means of which you think of it as a **substance** . . . Thus, convinced by the necessity with which this concept presses itself on you, you must concede that it has its seat in your faculty of cognition *a priori*" (*KrV* B6). It is through these kinds of transcendental arguments that Kant repeatedly seeks to establish the conclusion that certain components of our cognition are necessary and could not have come from experience.

It should be sufficiently clear at this point how Kant's method works in the case of theoretical principles. We can know that a concept (e.g., substance) has a pure a priori source when we discover that it is necessary for our representations and that it is not grounded empirically. Naturally these two are closely related, for anything which we judge as necessary cannot be justified by experience. And with this information in hand, we can return to the fact of reason in order to understand how our consciousness of the moral law passes this test.

3.2.3 The Necessity and Purity of the Moral Law

[158] Locke, employing a similar method, also finds that we posit substance as the thing "we know not what" which underlies the perceivable qualities of an object. He says "[S]ubstance is supposed always something besides the extension, figure, solidity, motion, thinking, or other observable ideas, though we know not what it is" (*Essay*, ch. XXIII).

The key to understanding the fact of reason is explaining how our consciousness of the moral law can live up to these two criteria of a priority (*viz*., necessity and non-empirical grounds) which were established in the first *Critique*. This is yet another rare instance where Kant makes illuminating use of examples.[159] Just before introducing the fact of reason, he presents one of his most important thought experiments:

> Suppose someone asserts of his lustful inclination that, when the desired object and the opportunity are present, it is quite irresistible to him; ask him whether, if a gallows were erected in front of the house where he finds this opportunity and he would be hanged on it immediately after gratifying his lust, he would not then control his inclination. One need not conjecture very long what he would reply. But ask him whether, if his prince demanded, on 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. He judges, therefore, that he can do something because he is aware that he ought to do it and cognizes freedom within him, which, without the moral law, would have remained unknown to him. (*KpV* 5:30)

In the first half of the experiment, we are supposed to learn the lesson that the drive of self-preservation is strong enough to motivate even a scoundrel to do what is right. The even more interesting point comes in the second half where Kant asks us to put that same scoundrel in a situation where the stakes are reversed. In the first situation, he was forced to comply with the moral law on pain of death, but now he must now choose between doing his moral duty and preserving his life. Kant thinks that the reader will immediately see how all the forces of inclination and self-love pull the man in the direction of bearing false witness, and, at the same time, he knows that he *ought* to act in accordance with the moral law.

Kant thinks that common, everyday moral experience provides ample evidence of this battle between the respect [*Achtung*] we have for the moral law and the incentive [*Triebfeder*] we

[159] Kant notoriously eschews the use of examples and he even tries to explain and justify the scarcity of illustrations in the first *Critique*. See A xviii–xix.

have to act on inclinations [*Neigung*] of self-love. Although these two forces sometimes coincide (e.g., the prudent shopkeeper who charges fair prices in order to maintain his reputation [*G* 4:397]), they can come apart as they do in the case of the prince asking for false testimony. Under such conditions, when all the inclinations of the empirical world implore an agent to violate the moral law, she remains aware that she ought to act in accordance with the moral law. Since the source of this unconditional reason cannot be empirical, Kant concludes that it must come from reason itself. He says that "it must be noted carefully that it [our consciousness of the moral law] is not an empirical fact but the sole fact of pure reason which, by it, announces itself as originally lawgiving" (*KpV* 5:31).

Later in the second *Critique*, Kant makes it clear that this is precisely the lesson that we are meant to draw from these thought experiments. He compares them to an experiment that a chemist might perform:

> He [a philosopher] has, however, the advantage that, almost like a chemist, he can at any time set up an experiment with every human practical reason in order to distinguish the moral (pure) determining ground from the empirical, namely, by adding the moral law (as a determining ground) to the empirically affected will (e.g., that of someone who would gladly lie because he can gain something by it). When an analyst adds alkali to a solution of calcareous earth in hydrochloric acid, the acid at once releases the lime and unites with the alkali, and the lime is precipitated. In just the same way, if a man who is otherwise honest (or who just this once puts himself only in thought in the place of an honest man) is confronted with the moral law in which he cognizes the worthlessness of a liar, his practical reason (in its judgment of what he ought to do) at once abandons the advantage, unites with what maintains in him respect for his own person (truthfulness), and the advantage, after it has been separated and washed from every particle of reason (which is altogether on the side of duty), is weighed by everyone, so that it can enter into combination with reason in other cases, only not where it could be opposed to the moral law, which reason never abandons but unites with most intimately. (*KpV* 5:92–93)

The thought experiment invites the reader to notice not only that morality and self-love can come apart but that they have utterly distinct origins and are fundamentally different in kind "as if the

two were unmixable chemical compounds."[160] Kant takes this to show that the moral law cannot have any empirical ground, since the moral law can generate reasons that act as a counterweight to all empirically grounded inclinations. Owen Ware sums this up well: "When we attend to the necessity of the moral law, we are led to see that it wholly excludes empirical motives from entering into our maxims. No other principle — hypothetical or prudential — has this unconditional character."[161]

Without this purity, the moral law would not have a special status. The categorical oughts that it purports to generate would be nothing more than hypothetical oughts which hold only contingently, and this would rob them of their worth. Kant makes it clear that this dignity is what hangs in the balance:

> From what has been said it is clear that all moral concepts have their seat and origin completely a priori in reason, and indeed in the most common reason just as in reason that is speculative in the highest degree; that they cannot be abstracted from any empirical and therefore merely contingent cognitions; that just in this purity of their origin lies their dignity, so that they can serve us as supreme practical principles; that in adding anything empirical to them one subtracts just that much from their genuine influence and from the unlimited worth of actions . . . (*G* 4:411)

Keeping in mind the way in which the fundamental worth and dignity of rational beings is the ground of human rights (as I explained in chapter 1), we can see how much of Kant's moral and political philosophy hinges on the claim that the moral law is not tainted by empirical grounds.

That is the sense in which the moral law is (and must be) pure. But what is the sense in which it is necessary? One possibility is that the necessity refers to the relationship that the moral law has to our will. As Kant says in the *Groundwork*, "the determination of such a will in conformity with objective laws is *necessitation* [*Nötigung*]." (4:413).[162] When we choose to act in

[160] Ware, "Rethinking Kant's Fact of Reason," 12.
[161] *Ibid*., 16.
[162] "[D]ie Bestimmung eines solchen Willens, objektiven Gesetzen gemäß, ist Nötigung."

conformity with the moral law, our will, which is subject to inclinations, is *necessitated* by the objective practical law. Although this necessitation looks like a viable candidate, this cannot be the sense of necessity at stake here.[163] The problem is that this relationship between the law and the will holds only for beings like us who are rational but who are "also subject to subjective conditions (certain incentives) that are not always in accord with the objective ones" (*G* 4:413). It was pointed out earlier that this relationship does not obtain for "a perfectly good will" which is also subject to the moral law but which "could not on this account be represented as *necessitated* to actions in conformity with the law since of itself, by its subjective constitution, it can be determined only through the representation of the good" (4:413). Necessitation [*Nötigung*] is Kant's technical term for the way in which the moral law constrains the *human* will and compels us to act against (or at least be indifferent to) our inclinations. Since purely rational beings (like angels or God) are not subject to inclinations, the moral law cannot be necessary for them in this way. So we must find another kind of necessity in order to avoid the accusation that the moral law's necessity holds only for human beings — a conclusion that Kant is at pains to avoid.[164]

Fortunately, this is not the sense in which the moral law is necessary. The necessity in question is the one that we find when we reflect on the difference between prudential oughts and moral oughts. Recall that one of the most notable refrains of Kant's moral philosophy is the importance of the distinction between hypothetical and categorical imperatives:

> For, only a law brings with it the concept of an unconditional and objective and hence universally valid necessity [*Notwendigkeit*], and commands are laws that must be obeyed, that is must be followed even against inclination. Giving counsel does not involve necessity, which, however, can hold only under a subjective and contingent condition, whether this man or that man counts this or that in his

[163] See Guyer, "Problems with Freedom," 192.

[164] As he says in several places, "[I]t is of the greatest practical importance not to make its principles dependent upon the special nature of human reason — as speculative philosophy permits and even at times finds necessary — but instead, just because moral laws are to hold for every rational being as such, to derive them from the universal concept of a rational being as such" (*G* 4:412–13).

> happiness; the categorical imperative, on the contrary, is limited by no condition and, as absolutely although practically necessary, can be called quite strictly a command . . .
>
> 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 [*Notwendigkeit*] 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. (*G* 4:416; 4:420–21).

The necessity that Kant refers to here is the absolute bindingness (for all rational beings, including non-human ones) of the unconditional ought. Unlike prudential oughts which command actions only insofar as some contingent end is desired, oughts generated by the categorical imperative are unconditionally necessary; they are recognized as objectively valid. The prescriptions of the moral law are not mere suggestions or counsels of prudence, they are requirements of rational agency. He echoes this point from the *Groundwork* in the second *Critique* just before his first mention of the fact of reason: "[T]he rule says: one ought absolutely to proceed in a certain way. The practical rule is therefore unconditional and so is represented a priori as a categorical practical proposition by which the will is objectively determined absolutely and immediately" (*KpV* 5:31).

As we have seen, the necessity of the moral law is linked to its purity. Our recognition of the necessity of the moral law indicates that it must also have a pure source. The thought experiment about the prince is supposed to *illustrate* the purity of the moral law, but it is not meant to serve as a deductive argument for that claim. We do not need to ask whether or not it is possible for pure reason to be practical; we know that it is possible because it is actual. As Kant puts it:

> But that pure reason, without the admixture of any empirical determining ground, is practical of itself alone: this one had to be able to show from the *most common practical use of reason,* by confirming the supreme practical principle as one that every natural human reason cognizes — a law completely a priori and independent of any sensible data — as the supreme law of its will. It was necessary first to establish and justify the purity of its origin even *in the judgment of this common reason* before science would take it in hand in order to make use of it, so to speak,

> as a fact [*Faktum*] that precedes all subtle reasoning about its possibility and all the consequences that may be drawn from it. (*KpV* 5:91)

Kant thinks that nothing can cast doubt on the possibility of pure practical reason because we are aware of its actuality and we are conscious of its activity. We cannot help but recognize the moral law and its necessity whenever we engage in practical deliberation.

To summarize, the first premise of the argument (that we are bound by the moral law) has now been fully explained and the fact of reason argument has been given as its justification. We are bound by the moral law in the sense that we have unconditional, overriding reasons for acting morally. And because of the objective necessity with which these reasons force themselves upon us, we can infer that they come from a non-empirical source, namely, pure reason. The painstaking exposition of these details will be indispensable in the next section where I will argue for the second premise of the argument: the moral law can bind our will in this way only if we possess transcendental freedom.

3.3 From the Moral Law to Transcendental Freedom

Some of the steps in the above arguments will once again serve as the premises for Kant's next move. In order to assert the reality of transcendental freedom, he must have successfully shown that *pure* reason can be practical. This is to be contrasted with the way that practical reason works when it operates in the interest of the sensibility. For instance, whenever we set an end for ourselves such as health or financial prosperity, practical reason is the faculty that determines the means through which we can achieve those ends. But in those instances, practical reason is the slave of the passions. Its ends are given externally; they do not come from within. Kant believes that he has shown that practical reason's capacities must go beyond this; it can generate its own

reasons for actions which do not depend on hypothetical ends — these are the moral oughts given a priori by a practical law.

The question now is how Kant plans to go from this conclusion — that pure reason generates unconditional oughts — to an assertion of the reality of transcendental freedom. Commentators have suggested a variety of strategies for making this move, but most of them are deeply problematic and fall short of justifying the move from the moral law to transcendental freedom. In section 3.3.1, I consider several possible arguments that I believe cannot ultimately succeed in justifying a belief in freedom. Most notably, I reject the widespread assumption that Kant's argument is nothing other than a simplistic version of the "ought implies can" principle. Many scholars assume that Kant's argument from the moral law to freedom does nothing more than move from the truth of the claim that we ought to follow the moral law to the claim that we have the freedom to act in such a way.

I suggest instead that Kant's argument is dependent on the existence of moral feeling — respect for the law. Since the fact of reason has been defined as the fact that pure reason generates unconditional oughts, this falls short of what Kant needs for demonstrating the reality of freedom. What is needed is now is the claim that these oughts can move us to act. Because these oughts have a pure source, if they can motivate action, then it must be granted that pure reason, independent of inclination, can indeed determine the will. I argue in section 3.3.2 that moral feeling is the bridge between the fact of reason and the assertion of freedom.

3.3.1 Common but Unsatisfying Arguments Connecting Morality and Freedom

Perhaps the most common interpretation of Kant's argument for freedom is that it boils down to the "ought implies can" principle. And, to be fair, this interpretation seems to have rather

abundant textual support. In *Religion Within the Boundaries of Mere Reason*, Kant invokes this principle at least seven times.[165] The first instance of it is in Part One, where Kant says that "However evil a human being has been right up to the moment of an impending free action . . . his duty to better himself was not just in the past: it is still his duty *now*; he must therefore be capable of it' (*Rel* 6:41). Or again, when discussing the idea of a Christ-like prototype which is pleasing to God, Kant says, "We ought to conform to it, and therefore we must be able to" (*Rel* 6:62).

One of Kant's main tasks in the *Religion* is to show how his moral theory is consistent with the fall and subsequently how conversion is possible (something I explain more fully in chapter 4). Since the noumenal realm is the abode of transcendental freedom and is thus an atemporal form of agency, it is difficult to understand how something like conversion is possible. The very idea of conversion seems to require that a will which was once corrupted was later transformed, and so temporality seems inextricably bound up with the concept of conversion. Wrestling with this difficult question, Kant resorts to the ought implies can principle as a way of dismissing this difficulty:

> How it is possible that a naturally evil human being should make himself into a good being surpasses every concept of ours. For how can an evil tree bear good fruit? But, since by our previous admission a tree which was (in its predisposition) originally good but did bring forth bad fruits, and since the fall from good into evil (if we seriously consider that evil originates from freedom) is no more comprehensible than the ascent from evil back to good, then the possibility of this last cannot be disputed. For in spite of that fall, *the command that we **ought** to become better human beings still resounds unabated in our souls; consequently, we must also be capable of it*.[166]

So, at least in this part of the *Religion*, Kant makes use of the principle. But it would be a mistake to think that Kant has recourse to it only in this context, where he faces the difficult task of trying to reconcile his moral philosophy with religious dogma. He invokes it in a variety of other contexts.

---

[165] There are more than a dozen references to it throughout Kant's writings.
[166] *Rel* 6:44–45.

As we saw in the gallows thought experiment, when the agent is asked to bear false witness against the honorable man, his awareness of the moral law allows him to cognize his freedom: "He judges, therefore, that he can do something because he is aware that he ought to do it and cognizes freedom within him, which, without the moral law, would have remained unknown to him" (*KpV* 5:30). The principle even makes an appearance in the first *Critique*:

> Pure reason thus contains — not in its speculative use, to be sure, but yet in a certain practical use, namely the moral use — principles of the possibility of experience, namely of those actions in conformity with moral precepts which could be encountered in the history of humankind. For since they command that these actions ought to happen, they must also be able to happen. (A807 / B835)

Kant's reasoning looks to be something like this:

> (1) Smith told a malicious lie.
> (2) If Smith does not have transcendental freedom, then he could not have done otherwise.
> (3) If Smith could not have done otherwise, then he was not morally obligated to tell the truth.
> (4) Smith was morally obligated to tell the truth.
> Therefore,
> (5) Smith has transcendental freedom.

The pressing question, of course, is what justifies the third premise (which is the contraposition of the ought implies can principle). Although Kant repeatedly invokes this principle, he never provides an argument for it, and this is disconcerting considering the weight that it must bear if it is to serve as the justification for transcendental freedom. To make matters worse, the above definition of the moral ought as an unconditional reason for action does not in any obvious way seem to entail the reality of transcendental freedom. There is no way that the mere existence of a reason for action (even one with a pure source) could justify a belief in the reality of freedom. For it is entirely possible that those reasons are causally impotent.

Spinozists, for instance, make a variety of ought claims in spite of their commitment to hard determinism. Sally the Spinozist would surely acknowledge that she *ought* not to have smoked a pack of cigarettes (since it made her passive and reduced her power to persevere in being), but granting the truth of this ought claim does not commit her to a belief in libertarian freedom. All that is contained in the Spinozistic ought is the idea that reason told her to act a certain way; it does not say anything about her *capacity* to have acted differently than she did in the actual world.

Pereboom argues that these weaknesses severely undermine Kant's argument for freedom, "[I]n our conception of morality, there are 'ought' judgments sufficient for morality that do not presuppose an 'ought implies can' principle; and second, there may be principles sufficient for morality that are not 'ought' judgments and are not undermined by an 'ought implies can' principle."[167] He is certainly right that a moral theory could employ principles that do not involve 'ought' claims (we could simply call actions 'good' and 'bad' and say nothing of 'ought'), but this possibility is off limits for Kant since the moral law confronts human beings in the form of 'oughts.' But his first objection poses a serious problem for anyone using 'ought implies can' to argue for freedom. Many ought claims do not imply the ability to do otherwise. A determinist might tell her daughter that she should not make false promises, and she does this as a way of guiding her daughter's actions. She believes that determinism is true, but she also believes that telling her daughter this might play a causal role in determining her future actions. Pereboom rightly concludes that "the truth of the action-guiding variety of 'ought' judgment is not affected by determinism and by agents' inability to have done otherwise."[168] In response, it could be said that Kant's 'oughts' are not simply action-guiding oughts; they are unconditional, overriding

[167] Pereboom, "Kant on Transcendental Freedom," 561.
[168] Ibid. 561

reasons for acting morally. But this still falls short of entailing that human beings have free will. Even if we have these kinds of reasons (and they spring from a pure source), we do not know whether or not these reasons can move us to act. *That* is what Kant needs to show. If moral oughts can indeed motivate actions (and have a pure source), then that would mean that the will can be determined autonomously by something which was generated spontaneously by a non-empirical faculty. The mere existence of moral oughts is not enough to satisfy the antecedent of that conditional.

Thus, the argument from the 'ought implies can' principle does not justify the assertion of transcendental freedom. There is another argument, which is similar to the first, that commentators sometimes attribute to Kant. According to this line of thinking, freedom can be inferred from our common moral understanding, especially attributions of praise and blame. In his discussion of the malicious lie in the first *Critique*, Kant says that even if we recognize the role that the laws of nature played in causally determining the man to lie (e.g., his bad upbringing, etc.), our attribution of blame is not thereby mitigated: "Now even if one believes the action to be determined by these causes, one nonetheless blames the agent" (A555/ B583). And since this blame requires transcendental freedom, Kant says that we must judge the action not in terms of the natural causes the determined the empirical character of the liar but in terms of his intelligible character which was free to do otherwise.

This argument is no better than the first, and there are several powerful objections to it. First, the argument uncritically assumes that we can legitimately blame the liar. Our blame might simply be unjustified. If transcendental freedom is indeed a necessary condition of legitimate blame and there is no such freedom, then we might be blaming someone who does not deserve it. Second, as Pereboom points out, even if we grant the imputation of blame as prima facie evidence,

it is much too weak to sustain the claims that are eventually based on it. In the *Metaphysics of Morals*, Kant asks us to consider a murderer who is the last prisoner on an island society about to be dissolved. He insists that the death penalty is the only suitable punishment, and that this conclusion arises out of our consideration of his desert. He is to blame for his wrongdoing. Kant says that the fittingness of the punishment "is shown by the fact that only by this is a sentence of death pronounced on every criminal in proportion to his *inner wickedness*" (6:333). But if we use a belief in freedom to justify doing serious harm to someone, then we surely need to set a high bar as our standard of evidence. Pereboom is right to criticize this justification. He says, "If one's justification for harmful behavior depended on the claim that we are transcendentally free, but we have little or no evidence for this claim, and the story we need to tell to reconcile transcendental freedom with our best empirical theories is barely credible, then that justification would be inadequate."[169] As I mentioned above, Allen Wood compared our attribution of freedom to a jury declaring someone innocent because there is a reasonable shadow of doubt. The defendant's innocence, like the possibility of freedom, cannot be utterly ruled out.[170] But even if Kant has shown that we cannot conclusively rule out the possibility of transcendental freedom, that is too low a bar to set as a justification for capital punishment. Pereboom says that "the guidance that belief in our transcendental freedom would provide is more aptly described as on the side of *prosecuting* attorney."[171] Given the stakes, the burden of proof is too high to just uncritically assume that moral blame is justified.

Additionally, there is a third line of criticism that could be leveled against the link between blame and freedom. In "Freedom and Resentment," P.F. Strawson criticizes the ordinary

[169] Pereboom, "Kant on Transcendental Freedom," 564.
[170] Wood, "Kant's Compatibilism," 83.
[171] Pereboom, "Kant on Transcendental Freedom," 564.

assumption of the connection between attitudes of blame and beliefs about freedom and determinism. On Strawson's account, our attitudes of resentment and our attributions of blame merely show "how much we actually mind, how much it matters to us, whether the actions of other people — and particularly *some* other people — reflect attitudes towards us of good will, affection, or esteem on the one hand or contempt, indifference, or malevolence on the other."[172] He argues that these reactive attitudes are a natural part of our way of life and that such an attitude "neither calls for nor permits, an external 'rational' justification."[173] This is notably distinct from the first objection which suggested that our attributions of blame might be mistaken or misguided. Strawson does not think that it is reasonable to adopt such an epistemic stance with respect to our reactive attitudes.

If any of these three objections goes through, then moral responsibility is not a sufficient reason for asserting the reality of transcendental freedom. It seems to me that all three of them cast doubt on the link between moral responsibility and freedom. The first and third objections point out that blame might constitute no evidence in favor of freedom, and the second argues that even if it does constitute some prima facie evidence, it is much too little for the task at hand.

Since moral responsibility and the 'ought implies can' principle fail to justify the assertion of transcendental freedom, we must continue our search for a sound argument unless we want concede that Kant has no good reason for asserting the reality of freedom. It bears repeating that Kant believes "a free will and a will under moral laws are one and the same" (*G* 4:447). As he puts it in the second *Critique*, "freedom and the unconditional practical law reciprocally imply each other" (*KpV* 5:29). The moral law is the *ratio cognoscendi* of freedom, but the other half of that famous line is that freedom is the "*ratio essendi* of the moral law" (*KpV* 5:5). Thus, on Kant's

[172] Strawson, "Freedom and Resentment," 5.
[173] Ibid. 23.

view, we come to know freedom through the moral law, but the moral law could not exist within us without freedom. "Were there no freedom, the moral law *would not be encountered* at all in ourselves" (*KpV* 5:5). And so it might seem natural to assume that the fact of reason alone provides sufficient evidence for the reality of transcendental freedom. After all, the fact shows that we do encounter morality in us, and he emphatically insists that freedom is a necessary condition of this discovery.

But this move is not as simple or unproblematic as it seems. First, we must note that in the first six sections of the second *Critique* Kant proceeds analytically. He is merely unpacking definitions. From conceptual analysis alone, he asks what pure practical reason is, what its law would have to be, and the nature of a will that could act according to this law. Immediately before his statement of the reciprocity thesis, he outlines the two halves of the biconditional. The conditional that moves from morality to freedom is explained as follows:

> 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 . . . [I]f no determining ground of the will other than that of 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. But such independence is called *freedom* in the strictest, that is, in the transcendental, sense. Therefore, a will for which the mere lawgiving form of a maxim can alone serve as a law is a free will. (*KpV* 5:28–29)

Up to this point in the *Critique*, nothing has been asserted other than what follows from basic definitions of pure practical reason and an analysis of practical laws and the will. The fact of reason follows the reciprocity thesis, and Kant then proceeds to argue for the reality of freedom. So it is natural to assume that the fact satisfies the antecedent of the conditional. But a closer inspection shows that this cannot be the case. What Kant says is that if lawgiving form alone is sufficient for determining the will (i.e., if pure reason can be practical), then the will is free. The fact of reason, as I explained above, does not guarantee the truth of this claim. The fact shows us only that pure

reason generates oughts; it says nothing of our capacity to act on the basis of those oughts. It does not amount to the claim that pure reason can be practical. Indeed, if it did, then Kant would surely be guilty of begging the most important question of the second *Critique.* Rather than arguing for the conclusion that pure reason can be practical or for the connection between morality and freedom, he would simply be assuming those things from the outset.

Kant requires more premises to make the connection between the fact of reason and freedom. It is not enough that pure reason provides us with oughts; it must also give rise to an incentive or interest. For according to Kant's theory of action, "An interest is that by which reason becomes practical, i.e., becomes a cause determining the will" (*G* 4:460n). I argue in the next section that this is precisely what Kant must show in order to connect the fact of reason with the claim that pure reason can be practical.

3.3.2 The Missing Link: Respect for the Law and Moral Feeling

The problem here can be stated as a dilemma. Either the fact of reason shows us that pure reason can be practical or it does not. If it does, then the second *Critique* begs the question that it sets out to answer. If it does not, then the fact of reason alone does not justify the move to freedom. In his set up of the reciprocity thesis, Kant claims that if pure reason can be practical (i.e., if "the mere lawgiving form of maxims is the only sufficient determining ground of a will" [5:29]) then we would have a practical proof of transcendental freedom. For, according to his theory of action, moral oughts come from pure reason and do not originate from (or take any influence from) the empirical world or its natural laws. So if these moral oughts are capable of determining the will to action (independent of any desired end), then that would indeed show that the will is free. The interpretation I have given takes the fact of reason to be nothing more than the claim that pure

reason gives us unconditional reasons to act a certain way; it says nothing of the capacity of these oughts to serve as the determining ground of the will. A mediating concept is needed to bridge the gap between the existence of these moral reasons and our capacity to act on them. In this section, I argue that moral feeling is the mediating concept that Kant needs.

Once moral feeling is added as a premise, the argument from the fact of reason to transcendental freedom takes on a new form:

1. There are moral oughts which are given by pure reason alone.
2. If (1), then we experience moral feeling only if pure reason determines the will.
3. We experience moral feeling.
4. If pure reason can determine the will, then the will has transcendental freedom.

Therefore,

5. The will has transcendental freedom.

The first premise is simply the fact of reason (as I have defined it), and the fourth premise is one half of the reciprocity thesis (i.e., if pure reason is practical, then the will is free in the transcendental sense [*KpV* 5:28–29]). If the truth of those premises has been granted, then the soundness of the argument depends only on the claim that we are subject to moral feeling and the conceptual claim pure reason's capacity to determine the will directly is a necessary condition of moral feeling. The task of this section is to explain and defend these premises. I begin by offering an interpretation of what moral feeling is and what role it plays in motivating moral behavior. This in turn requires some explanation of Kant's theory of action. We must develop a more thorough understanding of what maxims are and what role feelings, incentives, and interests play in order to understand the significance of moral feeling.

For years, Kant struggled to answer the following question: How can the judgment that an action is morally right, give rise to an incentive that moves the will. He put this point especially well in a lecture he gave in the mid-1780s:

> The moral feeling is a capacity for being affected by a moral judgment. When I judge by understanding that the action is morally good, I am still very far from doing this action of which I have so judged. But if this judgment moves me to do the action, that is the moral feeling. Nobody can or ever will comprehend how the understanding should have a motivating power; it can admittedly judge, but to give this judgment power so that it becomes a motive able to impel the will to performance of an action — to understand this is the philosophers' stone.[174]

Here Kant recognizes, more clearly than anywhere else in his lectures or published works, the gap between the intellectual judgment that an action is right and the incentive that must be present for the will to act on that judgment. Although he had come to recognize how crucial moral feeling is at this point in his thinking (around the time of the *Groundwork*), the concept is still shrouded in mystery. Kant did not yet have a very clear idea of what this feeling is or what role it plays in moral motivation.

By the time of the second *Critique*, however, Kant had worked out the details, and he devoted an entire chapter of the Analytic to the topic of moral feeling as an "incentive" [*Triebfeder*] and its role in giving rise to an "interest" [*Interesse*] in morality. He begins the chapter with a discussion of moral worth, and the treatment here is quite similar to the one given in the *Groundwork*:

> What is essential to any moral worth of actions is *that the moral law determine the will immediately.* If the determination of the will takes place *conformably* with the moral law but only by means of a feeling, of whatever kind, that has to be presupposed in order for the law to become a sufficient determining ground of the will, so that the action is not done *for the sake of the law*, then the action will contain *legality* indeed but not *morality.*[175]

Kant argued throughout the analytic that the moral law must determine the will directly and immediately. That is, it must not direct the will by means of some end (e.g., a pleasant feeling that one gets acting morally, fear of punishment by God, etc.); pure reason alone, without the assistance

---

[174] *Collins*; 27: 1428
[175] *KpV*; 5:71

of any feelings, must determine the will.[176] But when it comes to explaining exactly how this is possible, Kant admits that he cannot provide an answer:

> For, how a law can be of itself and immediately a determining ground of the will (though this is what is essential in all morality) is for human reason an insoluble problem and identical with that of how a free will is possible. (*KpV* 5:72)

This concession echoes the conclusion of *Groundwork* in which Kant says

> [I]t is quite beyond the capacity of any human reason to explain *how* pure reason, without other incentives that might be taken from elsewhere, can be of itself practical, that is, how the mere *principle of universal validity of all maxims as laws* (which would admittedly be the form of pure practical reason), without any matter (object) of the will in which one could take some interest in advance, can of itself furnish an incentive and produce an interest that would be called purely *moral*: it is impossible for us to explain, in other words, *how pure reason can be practical*, and all the pains and labor of seeking an explanation of it are lost. It is just the same as if I tried to fathom how freedom itself as a causality of the will is possible. For then I leave the philosophic ground of explanation behind and I have no other. (*G* 4:461)

These are both expressions of what I call the Incomprehensibility Thesis, which I discuss in Chapter 4. Given that we cannot comprehend how freedom is possible, we must begin by acknowledging that moral feeling lies downstream of this incomprehensible determination. Pure reason is (in some way that we cannot comprehend) capable of determining the will without any assistance from feelings or empirical incentives. Moral feeling follows from this determination as an effect. As Kant puts it, "What we shall have to show a priori is, therefore, not the ground from which the moral law in itself supplies an incentive but rather what it effects (or, to put it better, must effect) in the mind insofar as it is an incentive" (*KpV* 5:72). We cannot understand how pure reason determines the will directly, but we are capable of knowing the *effect* this determination has on our sensible desires and inclinations.

[176] This is what Kant means when he keeps repeating the claim that no material principle is suitable for morality and that the form of universal law alone should determine the will.

This effect manifests itself as an opposition to two self-regarding tendencies: self-love [*Eigenliebe*] and self-conceit [*Eigendünkel*]. An agent acts out of self-love whenever she does something for the sake of satisfying an inclination. Any action which is done in order to attain some end (e.g., secure wealth, feel pleasure, etc.) is done out of self-love. Self-conceit, by contrast, takes place only when the agent elevates the principle of self-love, treating self-interested inclinations as having an objectively valid force. Self-conceit is the condition that arises when "self-love makes itself lawgiving and the unconditional practical principle" (5:74). "Self-conceit involves an illegitimate promotion of self-love, our natural tendency to care for the inclinations, to a position of first rank in the will."[177] Moral feeling, according to Kant, first arises as a negative effect on self-love and self-conceit. He claims that "Pure practical reason merely *infringes upon* self-love, inasmuch as it only restricts it . . . But it *strikes down* self-conceit altogether, since all claims to esteem for oneself that precede accord with the moral law are null and quite unwarranted because certainty of a disposition in accord with this law is the first condition of any worth of a person" (5:73).

One way to understand these two effects is by an analogy with defeaters in epistemology. Epistemologists have drawn a distinction between undercutting and rebutting defeaters.[178] For example, Jones believes that there was a UFO flying above her house. Her evidence for this belief

[177] Ware, "Kant on Moral Sensibility," 735. Among all the recent commentaries on the subject of moral feeling, Ware's analysis stands out as one that is especially sensitive to the origin of self-conceit and the first-person perspective that we must adopt in order to understand the connection between humiliation and respect. When describing why we have the propensity to elevate self-love to the status of self-conceit, he says, "we are prone to mistake a maxim of satisfying the inclinations for an unconditional principle of the will. Our subsequent 'act' of conferring authority onto sensibility is tempting, and naturally so, because it rests on an illusion that arises from the perspective we have as rational beings with needs and inclinations. We think the claims of our sensible nature should have priority in deliberation because they are the first in time to make demands on us, and they make demands on us repeatedly (we always have desires, even after developing our rational capacities). Consequently, we are liable to think that pursuing "happiness" (*Glückseligkeit*), the sum-total satisfaction of our desires, should take priority in all matters of choice" (736).

[178] See, for instance, Pollock, *Contemporary Theories of Knowledge*.

is that she saw a flying saucer at 2 PM on Saturday. Her neighbors could offer evidence against this belief by saying that they were also outside at 2 PM on Saturday, and they did not see any such object. This would be a rebutting defeater for Jones's belief. If, however, Smith tells Jones that he gave her a powerful hallucinogenic drug at 1 PM, that would be an undercutting defeater. The difference is that Smith is giving Jones a reason to doubt the connection between her evidence (her eyewitness account) and her belief (that there was a flying saucer above her house). The neighbors did not try to undercut the connection between Jones's evidence and her belief, they simply offered new evidence for the opposite belief.

So it is with self-love and self-conceit. The moral law infringes on inclinations in that it gives the agent a reason to act contrary to the dictates of inclination. Moral reasons are a counterweight to the incentives given by inclinations as the neighbors' observation (or lack thereof) was a counterweight to Jones's evidence. The moral law's relation to self-conceit is quite different. Rather than merely offering a counterweight, it "*strikes down*" self-conceit. That is, it undercuts self-conceit's mistaken claim of legitimacy that is "arrogantly"[179] given to self-love.[180] Self-conceit elevates self-love by treating it as if it were worthy of our respect — worthy of taking pride of place in our will — but our consciousness of the categorical imperative nullifies this claim of legitimacy, since the moral law alone, with its rationally authoritative imperative, is worthy of

[179] Kant uses the word '*Arrogantia*' to describe self-conceit. See *KpV*; 5:73.

[180] To fill out the details of the analogy, inclinations of self-love tells the agent to Φ, and the moral law tells her to not-Φ. These two opposing forces compete with one another and thus the moral law infringes on self-love. Similarly, Jones evidence leads her to believe P, but her neighbors' evidence leads her to believe not-P. In this way, Jones must choose what to believe given these two competing pieces of evidence. Self-conceit, however, is the propensity to prioritize self-love as if it had the force of law. Consciousness of the moral law undermines this claim of legitimacy by showing that it rests on an illusion and that the moral law alone can serve as a practical law. The role of illusion in the case of self-conceit is analogous to the undercutting defeater in the sense that the information about the hallucinogenic drug does not simply give Jones reason to believe not-P, it severs the connection between her evidence and her belief by showing her that her evidence was based on an illusion (hallucination).

such elevation. Thus, Kant says that the moral law brings about the "humiliation" (*Demütigung*) of self-conceit.

Kant claims that the feeling that results from this experience is a painful one. This feeling of pain is especially significant, for this is the first intersection (within the domain of practical reason) between pure practical reason and the sensibility:

> [T]he effect of the moral law as incentive is only negative, and as such this incentive can be cognized a priori. For, all inclination and every sensible impulse is based on feeling, and the negative effect on feeling (by the infringement upon the inclinations that takes place) is itself feeling. Hence we can see a priori that the moral law, as the determining ground of the will, must by thwarting all our inclinations produce a feeling that can be called pain; and here we have the first and perhaps the only case in which we can determine a priori from concepts the relation of a cognition (here the cognition of a pure practical reason) to the feeling of pleasure or displeasure. (*KpV* 5:72)

This is the first glimmer of hope in his search for the philosophers' stone; here for the first time, Kant believes he has shown that a judgment of pure reason can (and indeed must) have palpable effects in the sensible world.

As usual, the negative component has a positive counterpart. The positive feeling that coincides with the negative feeling of pain (via infringement and humiliation) is respect (*Achtung*):

> So the moral law strikes down self-conceit. But since this law is still something in itself positive — namely the form of an intellectual causality, that is, of freedom — it is at the same time an object of *respect* inasmuch as, in opposition to its subjective antagonist, namely the inclinations in us, it *weakens* self-conceit; and inasmuch as it even *strikes down* self-conceit, that is, humiliates it, it is an object of the greatest *respect* and so too the ground of a positive feeling that is not of empirical origin and is cognized a priori . . .
>
> Now, what in our own judgment infringes upon our self-conceit humiliates. Hence the moral law unavoidably humiliates every human being when he compares with it the sensible propensity of his nature. If something represented *as a determining ground of our* will humiliates us in our self consciousness, it awakens respect for itself insofar as it is positive and a determining ground. Therefore the moral law is even subjectively a ground of respect. (*KpV* 5:73)

Out of the negative aspect (the striking down of self-conceit), a positive ground is uncovered. We find a feeling of respect, which unlike any other feeling that we experience, does not have an empirical source. It comes from pure reason, and it is self-authored. The striking down of self-conceit is not the same as being chastised by a parent; the agent realizes that this censure comes from within. "Since this constraint is exercised only by the lawgiving of his *own* reason, it also contains something *elevating*, and the subjective effect on feeling, inasmuch as pure practical reason is the sole cause of it, can thus be called *self-approbation*" (5:80–81). The authority of reason is recognized in such a way that it makes the lure of inclination pale in comparison and seem small.

In this way, the feeling of respect is analogous to that of the sublime. In the third *Critique*, Kant describes the sublime as that "in comparison with which everything else is small" (*KU*; 5:520). In the case of the dynamical sublime, we perceive some force in nature so powerful and vast that our capacity to resist it cannot seem anything but small:

> Bold, overhanging, as it were threatening cliffs, thunder clouds towering up into the heavens, bringing with them flashes of lightning and crashes of thunder, volcanoes with their all-destroying violence, hurricanes with the devastation they leave behind, the boundless ocean set into a rage, a lofty waterfall on a mighty river, etc., *make our capacity to resist into an insignificant trifle in comparison with their power*. (*KU*; 5:261; emphasis added).

Following Burke, Kant thinks that the feeling of the sublime is not strictly pleasant. Kant says it is first felt like a fear or pain, and he quotes Burke saying that the feeling is "not, to be sure, pleasure, but a kind of pleasing horror, a certain tranquility that is mixed with terror."[181] The terrifying power

[181] *KU*; 5:277. Kant is quoting Burke's *A Philosophical Enquiry into the Origin of Our Ideas of the Sublime and Beautiful*, which greatly influenced Kant's aesthetics. Here is the quotation in its entirety: "[T]hat the feeling of the sublime is grounded on the drive to self-preservation and on **fear**, i.e., a pain, which, since it does not go as far as the actual destruction of bodily parts, produces movements which, since they cleanse the finer or cruder vessels of dangerous and burdensome stoppages, are capable of arousing agreeable sensations, not, to be sure, pleasure, but a kind of pleasing horror, a certain tranquility that is mixed with terror."

of nature makes our capacity to resist seem small, but, at the same time, whenever we experience the sublime, we must observe these natural phenomena from a place of safety. And, because of this, Kant thinks that we feel a sense of our "superiority over nature" as we take note of the fact that reason endows us with a power that subordinates the entirety of nature.[182] We thus experience the sublime as "a pleasure that is possible only by means of a displeasure" (*KU*; 5:260). The unpleasant feeling is our fear (in the case of the dynamical sublime) or our inadequacy (in the case of the mathematical sublime), but a pleasant sensation arises out of these feelings because reason asserts its superiority over nature.

That is just like what happens when our consciousness of the moral law humiliates our self-conceit and infringes on our self-love. Although this negative effect is itself a kind of pain, it alerts us to the presence of something sublime that resides in us as rational beings. So although it humiliates our self-conceit and makes our inclinations seem small, it elevates reason and esteems something within us. Unsurprisingly, Kant took notice of the striking parallelism between the moral feeling of respect and the aesthetic feeling of the sublime. In *Religion within the Bounds of Mere Reason*,[183] Kant makes the comparison:

> The majesty of the law instills awe (not dread, which repels; and also not fascination which invites familiarity); and this awe rouses the respect of the subject toward his master, except that in this case, since the master lies in us, it rouses a *feeling of the sublimity* of our own vocation and enraptures us more than any beauty. (*Rel*; 6:23)

[182] He says, "[T]he irresistibility of its power certainly makes us, considered as natural beings, recognize our physical powerlessness, but at the same time it reveals a capacity for judging ourselves as independent of it and a superiority over nature on which is grounded a self-preservation of quite another kind than that which can be threatened and endangered by nature outside us, whereby the humanity in our person remains undemeaned even though the human being must submit to that dominion" (*KU*; 5:261–62). The superiority comes from our power of reason. When it comes to the mathematical sublime, we perceive something so enormous and overwhelming that the imagination fails to comprehend it. But reason steps in and allows us to think of the infinite as a whole.

[183] Naturally, moral feeling is also a prominent theme in the *Critique of Judgment*. The phrase "moral feeling" appears more than a dozen times there, and Kant argues that cultivating taste in the context of appreciating art and nature plays an important role in cultivating our moral feeling (and vice versa).

Out of our negative feelings of infringement and humiliation, the feeling of respect emerges as "something elevating" in that we cognize our freedom as autonomous legislation, and this leads to a feeling of "self-approbation" [*Selbstbilligung*] (5:80–81).[184]

The final piece of the story has to do with the role that these incentives play in Kant's theory of action. Many commentators are rightfully wary of a "conflict of forces" interpretation. According to this reading, the feeling of respect and inclinations of self-love compete with one another, and whichever one is stronger ends up determining the will. In his critique of this interpretation, Allison describes the "conflict-of-forces" as the view that Kant "conceives of the moral life as essential one of conflict between psychic forces in which the human will is the playing field and prize rather than autonomous arbitrator. On this reading, it is by weakening the opposing forces of inclination (through a process of humiliation) that the feeling of respect prepares the way for the ultimate victory of the moral law as the superior force."[185] This conflict-of-forces view is notoriously problematic. The main problem is that it is at odds with Kant's view of how the will acts on maxims and how maxims incorporate incentives. On Kant's view, "incentives [*Triebfedern*] do not motivate by themselves causing action but rather by being taken as reasons and incorporated into maxims."[186] Thus, the "incentive (or motive) is denied any causal efficacy

---

[184] In his introduction to his translation of the third *Critique*, Paul Guyer also points out this similarity. He writes, "And this is what leads to the special quality of the experience of the sublime: unlike the experience of beauty, **it is not an unalloyed pleasure, but a complex feeling**, consisting first of frustration at the inability of the understanding to grasp an absolute whole with the assistance of the imagination, followed by pleasure at the realization of the fact that our imagination also reflects the demands of our reason (§ 27). This **complexity of the feeling of the sublime is akin to the complexity of the moral feeling of respect**, and leads Kant to the discussion of the dynamical sublime . . . The experience of the dynamical sublime is produced by the experience of vast forces in nature, such as those of towering seas or mountain ranges, in relation to which we realize that our own physical powers are puny. At the same time, however, the experience of our insignificance in relation to such physical forces also leads us to the realization that there is another force in us, the faculty of practical reason and the freedom of the will that it gives us, which gives us a value that cannot be damaged even by forces which would suffice for our physical destruction. **This again produces a complex mix of displeasure and pleasure, which is even closer to the moral feeling of respect**" (xxxi; emphasis added).

[185] Allison, *Kant's Theory of Freedom*, 126.

[186] *Ibid*. 51

apart from the adoption of a maxim by an agent to act on the basis of that incentive . . . Moreover . . . this act of adoption or incorporation is not itself causally conditioned'' (189). The will is not a passive faculty that acts on whichever incentive happens to be stronger; it freely chooses which incentives to incorporate into its maxim. And this certainly seems to be Kant's considered view by the time he writes *Religion within the Bounds of Mere Reason*:

> [F]reedom 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); only in this way can an incentive, whatever it may be, coexist with the absolute spontaneity of the power of choice (freedom)." (*Rel*; 6:23–24)

And although this quotation from the *Religion* is the one most frequently cited in support of the incorporation view, it is certainly present the second *Critique* as well. Kant explains how incentives become interests and how these are incorporated into maxims:

> Therefore respect for the moral law must be regarded as also a positive though indirect effect of the moral law on feeling insofar as the law weakens the hindering influence of the inclinations by humiliating self-conceit, and must therefore be regarded as a subjective ground of activity- that is, as the incentive to compliance with the law — and as the ground for maxims of a course of life in conformity with it. From the concept of an incentive [*Triebfeder*] arises that of an *interest* [*Interesse*], which can never be attributed to any being unless it has reason and which signifies an *incentive* of the will insofar as it is *represented by reason.* Since in a morally good will the law itself must be the incentive, the *moral interest* is a pure sense-free interest of practical reason alone. On the concept of an interest is based that of a *maxim.* A maxim is therefore morally genuine only if it rests solely on the interest one takes in compliance with the law.

Incentives constitute interests when reason represents them, and only then can they be incorporated into maxims. Allison is right to say that incentives can become interests only when reason "spontaneously *takes* an interest."

It is important to remember that all of these effects on the sensibility (infringement, humiliation, and respect) occur downstream[187] of the determining ground of the will. Kant warned at the outset of the chapter that moral worth requires pure practical reason to determine the will *immediately*; it must not determine the will by means of any mediating feeling. Thus, Kant's theory of moral action goes like this:

> First, the moral law determines the will objectively and immediately in the judgment of reason; but freedom, the causality of which is determinable only through the law, consists in just this: that it restricts all inclinations, and consequently the esteem of the person himself, to the condition of compliance with its pure law. This restriction now has an effect on feeling and produces the feeling of displeasure which can be cognized a priori from the moral law. (*KpV* 5:78)

The first step is that the moral law determines the will directly and immediately. Second, there is an element of displeasure (infringement on self-love, humiliation of self-conceit). Third, there is the recognition that "this humiliation . . . is an elevation of the moral — that is, the practical — esteem for the law on the intellectual side; in a word it is respect for the law, and so also a feeling that is positive in its intellectual cause, which is known a priori." In the third step, Kant cannot refrain from mentioning that freedom is implicated in that there is an intellectual cause. Indeed, freedom has been part of the process all along. The respect we have for reason (in our person and in others) is inextricably bound up with freedom. The effects of freedom are apparent throughout this story. Without freedom, how could pure reason have an effect on the sensible world? How could any feelings of pleasure or displeasure have a pure, non-empirical source? Freedom is in the background as a necessary condition of Kant's analysis of moral feeling. If such feelings are brought about by pure reason, then this proves that the moral law is the determining ground of the will. This is the final move in the argument from the fact of reason to transcendental freedom. Kant

[187] I have adopted the term 'downstream' instead of saying that these effects take place '*after*' the law is the determining ground of the will because of the implications that come with invoking temporality.

says, “The consciousness of a *free* submission of the will to the law, as yet combined with an unavoidable constraint put on all inclinations through only by one’s own reason, is respect for the law” (*KpV* 5:80). Freedom is a condition of the possibility of respect for the moral law. We cognize moral feeling as the effect that the moral law has when it determines the will by routing inclinations of self-love and undermining their claim on us. The moral law can have this effect only if pure reason is practical, and transcendental freedom is a necessary condition of pure reason having this capacity.

At the close of the chapter, Kant acknowledges that freedom is the only way to make sense of his account of moral interest, moral feeling, and the motive of duty. He says that duty with its “sublime and mighty name embraces nothing charming or insinuating but requires submission.” It commands us without the help of any enticing inclinations or any fear of punishment. It is a “law before which all inclinations are dumb, even though they secretly work against it,” and this leads him to ask “what origin is there worthy” of such a thing and how duty could nobly descend from the heavens and force us to reject “kinship with the inclinations.” His answer is that it can be nothing other than “freedom and independence from the mechanism of the whole of nature” (*KpV* 5:86–87). Only if we are intelligible beings with complete independence from the laws of nature could we be capable of this kind of self-legislation. Transcendental freedom alone makes room for the capacity to give ourselves a law of pure practical reason that directly determines our will, infringes on our inclinations, humiliates our self-conceit, and gives rise to an object of utmost respect.

As I explained in chapter 1, it is transcendental freedom which grounds morality and the dignity of human beings. I then argued in chapter 2 that theoretical reason was not able to show that we have this capacity or even that such freedom is metaphysically possible. In this chapter, I

presented Kant's practical argument for the conclusion that we do indeed have transcendental freedom. In the next chapter, I explain what it is about freedom that we cannot comprehend, what questions must be left unanswered, and why the incomprehensibility of freedom is an important part of Kant's view.

Chapter 4
The Incomprehensibility of Freedom

4.1 Introduction

At this point, it may seem like the important issues surrounding transcendental freedom have been addressed. Theoretical reason provided us with the concept (but left us agnostic about its status); practical reason showed us why we should believe in it. What more needs to be said? Indeed, many commentators write as if there is nothing left to say about transcendental freedom other than what I have already discussed in the first three chapters. But by treating Kant's view in such a way, they ignore what I believe to be one of its most important components: the incomprehensibility of freedom.[188]

In addition to the fact that theoretical reason cannot provide us with reason to believe (or disbelieve) in transcendental freedom, the epistemic boundaries of the first *Critique* also place limitations on how much we can learn about freedom after practical cognition compels us to believe in its existence. The nature of these limitations, Kant's argument for them, and their impact on his view have been unjustly neglected in the secondary literature. My objective here is to rectify this by showing what Kant's position is, how he argues for it, and what its implications are. The structure of the chapter is as follows. In the next section, I explain Kant's "incomprehensibility thesis" — the claim that the conditions of cognition give rise to epistemic limitations which make

[188] In-depth discussions of the incomprehensibility of freedom are very rare in the secondary literature. It is often mentioned only in passing as if it were something trivial or anodyne, not worthy of further investigation. I think this is a mistake, as I believe there are many lessons to be learned from a close analysis of the issue. To the best of my knowledge, the only work dedicated entirely to the incomprehensibility of freedom is Klaus Konhardt, "Die Unbegreiflichkeit der Freiheit. Überlegungen zu Kants Lehre vom Bösen." But the scope of Konhardt's paper is largely restricted to the concept of evil, which is only one of the components of the incomprehensibility thesis. Perhaps the most noteworthy exception to the general neglect of the incomprehensibility thesis is Allen Wood's work, which shows a deep appreciation of the issue. See Wood, *Kantian Ethics*, 137 and "Kant's Compatibilism," 99. I worry that Wood takes it too far, however, as he treats Kant's talk of transcendental freedom as fiction. I explain how my view diverges from his in section 4.4.

it impossible for us to grasp all the details of transcendental freedom. In particular, we cannot comprehend (1) the precise relation between intelligible character and empirical character or (2) the nature of the intelligible self or its free agency as a spontaneous first cause. I present a novel reconstruction of Kant's argument for this position. In section 4.3, I outline some of the negative implications of this conclusion, highlighting questions and debates which, if I am right about the incomprehensibility of freedom, will turn out to have no definitive answer. I then turn to the positive implications, explaining Kant's view that freedom creates value *sui generis* and why we are better off acknowledging its incomprehensibility. Although it might seem disappointing at first, I think that this facet of reason's "peculiar fate" should ultimately be seen in a positive light. I conclude in section 4.4 by considering some objections and offering replies.

4.2 The Incomprehensibility Thesis

My aim in this section is threefold. First, I present textual evidence showing that Kant is committed to the idea that certain things about transcendental freedom are incomprehensible. My second aim is to explain exactly what it is about freedom that Kant thinks is beyond our comprehension. He surely does not mean that we cannot comprehend *anything* about freedom. As I showed in chapter 2, we must be able to understand how it is logically consistent with his commitment to natural necessity.[189] I will therefore distinguish between the things that Kant thinks we can and cannot know about freedom. My third goal is to argue that he has principled reasons for the incomprehensibility thesis. It might be tempting to think that the incomprehensibility of freedom is a last resort, a subterfuge Kant employs after failing to give a satisfactory theoretical explanation for something that his practical philosophy requires us to believe. I think we should

[189] See A445–51/ B473–79; A531–58/ B559–86

resist that temptation by coming to see how the incomprehensibility thesis follows directly from Kant's epistemological commitments. I have divided this section into two parts. First, I show what Kant said about the incomprehensibility of freedom, highlighting the most salient points. I identify what I take to be the four principal themes of Kant's comments on the incomprehensibility of freedom. These pertain to both the content of and the reasoning behind the Kant's claim. I elaborate on these points in section 4.2.2 where I synthesize them into an argument for the incomprehensibility thesis.

4.2.1 Kant on the Incomprehensibility Thesis

There can be little doubt about Kant's commitment to the incomprehensibility of freedom; the textual evidence is overwhelming. In his published works alone, it shows up more than twenty times.[190] Here are three of the most perspicuous instances from the *Groundwork*:

> **(1)** Philosophy must therefore assume that no true contradiction will be found between freedom and natural necessity in the very same human actions, for it cannot give up the concept of nature any more than that of freedom. Nevertheless, this seeming contradiction must be removed in a convincing way, *even though we shall never be able to comprehend how freedom is possible*" (4:456; emphasis added).

> **(2)** But reason would overstep all its bounds if it took it upon itself to *explain how* pure reason can be practical which would be exactly the same task as to explain *how freedom* is possible.
>
> For we can explain nothing but what we can reduce to laws the object of which can be given in some possible experience. Freedom, however, is a mere idea, the objective reality of which can in no way be presented in any possible experience; and because no example of anything analogous can ever be put under it, it can never be comprehended or even only seen. (*G* 4:458–59)

[190] Twenty might even be a conservative number. It can be found in the following places: *G* 4:456, 4:459, 4:460, 4:461, 4:462; *KpV* 5:7, 5:45–49, 5:72, 5:94, 5:99, 5:133, *MS* 6:418; *KrV* A448/B476, A557/B585, *Rel.* 6:138, 6:145, 6:171, 6:191; and *KU* 5:275. It also pops up in various *Reflexionen* dating back as far the 1760s, and it appears in both metaphysics and ethics lectures. I quote some of these below, and I have numbered all the quotations containing the incomprehensibility thesis.

> **(3)** But it is quite beyond the capacity of any human reason to explain *how* pure reason, without any other incentives that might be taken from elsewhere, can be of itself practical, that is, how the mere *principle of the universal validity of all its maxims as laws* (which would admittedly be the pure form of a pure practical reason), without any matter (object) of the will in which one could take some interest in advance, can of itself furnish an incentive and produce an interest that would be called purely *moral*; it is impossible for us to explain in other words, *how pure reason could be practical*, and all the pains and labors of seeking an explanation of it are lost.
>
> It is just the same as if I tried to fathom how freedom itself as the causality of a will is possible. For then I leave the philosophic ground of explanation behind and I have no other. I might indeed revel in the intelligible world, the world of intelligences, which is still left to me; but even though I have an *idea* of it, which has its good grounds, yet I have not the least *cognizance* of it nor can I ever attain this by all the efforts of my natural faculty of reason. (*G* 4:461–62)

These passages from *Groundwork* III contain several illuminating clues about both the meaning of and the reasoning behind the incomprehensibility thesis. First, it is clear that the conditions of cognition are part of the story. The intelligible self, which possesses transcendental freedom, resides in the noumenal realm, placing it beyond the limits of possible experience and cognition. And as Kant says here, "[W]e can explain nothing but what we can reduce to laws the object of which can be given in some possible experience." This partially explains the incomprehensibility. We have an idea (*Idee*) of the intelligible self as a free cause, but we have no cognizance (*Kenntnis*) of it. This tracks a distinction that Kant, which I discussed in chapter 2: the difference between our ability to think (*denken*) of something and to cognize it (*erkennen*).[191] When it comes to thinking about freedom of the intelligible self, no object is represented through sensible intuition because nothing is given to us except through the forms of sensibility — space and time. The intelligible self, however, is not in space and time. We are well aware by now that being outside of space and time is the very thing that exempts it from the universal causal necessity of the empirical world. Thus, when we think about transcendental freedom, we are not dealing with a singular, immediate

[191] Kant is especially clear about delineating thought and cognition in the second edition. See B146 and Bxxvi.

representation of an object given through intuition. It is a mere idea which "signifies only a 'something' that is left over when I have excluded from the determining grounds of my will everything belonging to the world of sense" (*G* 4:462). We can revel in thoughts of the intelligible world, but when we do, we fall short of comprehension as we are no longer on the terra firma of cognition.[192]

Furthermore, transcendental freedom requires "an **absolute** causal **spontaneity**" — a causality "through which something happens without its cause being further determined by another previous cause" (A446/ B474). By contrast, our cognition of objects of appearance necessarily involves a causal nexus; nothing happens unless it was determined by a previous state "in accordance with the law of the connection of cause and effect" (B232). We cannot represent transcendental freedom by means of a singular, immediate representation because the conditions of cognition are too restrictive. Cognition requires that an object be given, and no object is given in this case.[193] Thoughts are not limited in this way. Kant says, "I can think whatever I like, as long as I do not contradict myself i.e., as long as my concept is a possible thought, even if I cannot give

---

[192] The repetition of the word 'comprehend' throughout the dissertation (and particularly this chapter) is burdensome but necessary. 'Comprehend' is a technical term for Kant, and he uses it consistently in the context of freedom (he gives both the German '*begreifen*' and the Latin '*comprehendere*'). In the *Jäsche Logic*, he distinguishes between seven grades of cognition, and comprehension is the highest. For a detailed discussion of cognition, including Kant's seven-fold distinction in the *Jäsche Logic*, see Watkins and Willaschek, "Kant's Account of Cognition."

Most important for my purposes here: "The seventh, finally: to comprehend [*begreifen*] something (*comprehendere*), i.e., to cognize something through reason or a priori to the degree that is sufficient for our purpose. For all our comprehension is only relative, i.e., sufficient for a certain purpose; we do not comprehend anything without qualification" (*LJ* 9:64–65). In the *Vienna Logic*, Kant provides the example of our ability to comprehend the path of the moon sufficiently to predict a lunar eclipse (24:846). We can comprehend the moon's orbit and make such predictions through reason without having to wait and see whether or not an eclipse will occur on a particular day.

Thus, when Kant says that "we will never comprehend how freedom is possible [*wie Freiheit möglich sei, niemals begreifen Könnte]"* (*G* 4:456) or when he refers to the "incomprehensibility [*Unbegreiflichkeit*]" (*KpV* 5:7) of freedom, this technical definition is what he has in mind. We have no insight a priori through reason into freedom which is sufficient for *theoretical purposes*. As I argue below in the context of the **theoretical knowledge constraint**, our comprehension of freedom is perfectly sufficient for practical purposes but not theoretical purposes. See **(4)**, **(5)**, **(6)** and *Rel* 6:145.

[193] As I explain below, failing to have determinate content given through intuition is only one of two reasons why we cannot have theoretical cognition of a spontaneous first cause. The second reason is that our empirical, schematized concept of causality, as Kant explains it in the Second Analogy, would not able to process such a cause.

any assurance whether or not there is a corresponding object somewhere within the sum total of all possibilities" (Bxxvi). In the case of freedom, we merely apply the pure (i.e., unschematized) concept of causality to the idea intelligible self, thinking of it as the ground of certain appearances in the empirical world. But this does not constitute a theoretical cognition. And that places a limitation on our knowledge and comprehension of freedom. This is the first recurrent theme to note about the incomprehensibility thesis. I will refer to it as the **cognition constraint**: our ability to comprehend freedom is constrained by the conditions of cognition.[194]

Second, in the above excerpts from the *Groundwork*, Kant says that explaining "how pure reason can be practical" is "just the same as" explaining how freedom is possible; it is "exactly the same task." This gives us some additional information about the content of the incomprehensibility thesis. To see why freedom is incomprehensible, we need to understand what Kant means when he says that we cannot explain how pure reason can be practical. In short, the issue is that we cannot comprehend how morality could interest us at all: "The subjective impossibility of *explaining* the freedom of the will is the same as the impossibility of discovering and making comprehensible an *interest* which the human being can take in moral laws . . ." (*G* 4:460). To put it in Kant's terms, we cannot understand how the pure form of universal law, without the help of empirical incentives (*Triebfedern*), can produce an interest (*Interesse*) in us and serve as the determining ground of the will.[195] Explaining this requires us to revisit Kant's theory of agency.

---

[194] In this section, I identify four central themes of the incomprehensibility thesis; I call two of them "constraints" and the other two "problems." I think of the problems as components of transcendental freedom that we cannot comprehend. The "constraints" by contrast are foreclosed possibilities of comprehending freedom. They pertain to means of cognitive access, which we do not possess, that would allow us to comprehend freedom.

[195] These terms, which I mentioned several times in the preceding chapters, are defined in various places throughout Kant's writings. In the *Groundwork*, Kant says that an incentive is "the subjective ground of desire" (*G* 4:428). In the second *Critique*, he says that an incentive is "the subjective determining ground of the will of a being whose reason does not by its nature necessarily conform with the objective law" (5:72). And he says that an interest "signifies an *incentive* of the will insofar as it is *represented by reason*" (*KpV* 5:79). Cf. *MS* 6:212, *G* 4:414n. For more on incentives, interests, and Kant's moral psychology, see Josefine Nauckhoff, "Incentives and Interests in

Recall (from chapters 1 and 3) that when we take something as a reason for action, we generally have an interest in the pursuit of some end (e.g., money, good health, approval from others, etc.). In these instances, we are driven by an empirical incentive. These ends are at the heart of most hypothetical imperatives. For example: "He must work and save in his youth in order to not want in his old age" (*KpV* 5:20). In such cases, the "matter of the practical principle" (i.e., the desired end) is the determining ground of the will. Thus, the prescription has a normative force only insofar as the end is desired. The young man could dispense with this imperative if he "does not hope to live to old age, or thinks that in case of future need he can make do with little" (*KpV* 5:20). It is precisely because of this contingency that Kant argues that practical principles grounded in the desire of some object, "can furnish no practical laws" (*KpV* 5:21). Rather than acting in the pursuit of some end, acting out of respect for the moral law requires us to make the "form of law" the determining ground of the will. But we are unable to comprehend how an idea of pure reason, something that is entirely severed from the empirical world, could generate an interest in us. How can we take an interest in a judgment of pure reason — something devoid of empirical incentives?

Kant has additional reasons for finding this troubling because he believes it is impossible for us to comprehend how an idea of reason could cause pleasure or displeasure. We cannot understand how a pure faculty, could have an effect on sensibility. The mechanism behind moral feeling is shrouded in mystery. He says that it is "quite impossible to see, that is, to make comprehensible a priori, how a mere thought which contains nothing sensible produces a feeling of pleasure or displeasure; for that is a special kind of causality about which, as about any causality, we can determine nothing whatever a priori but must consult experience alone" (*G* 4:460). Our empirical concept of causality can connect only sensible objects with one another. Such knowledge

---

Kant's Moral Psychology"; Patrick Frierson, *Kant's Empirical Psychology*; Stephen Engstrom, "The *Triebfeder* of Pure Practical Reason"; and Herrera, "Moral *Triebfeder*."

must be gained through experience. This problem case (pure reason's effect on sensibility) involves a causal connection between a pure idea, which does not originate in the sensible world, and a sensible feeling of pleasure or displeasure. There is no experience that could give us insight into how this works.

The intelligible world contains no trace of incentives from the world of sense. When we try to think of pure reason "with reference to a pure world of understanding as a possible efficient cause" we inevitably discover that "an incentive must be quite lacking" (*G* 4:462). Kant suggests that we might then suppose that the idea of the intelligible world *itself* "would have to be the incentive or that in which reason originally takes an interest; *but to make this comprehensible is precisely the problem that we cannot solve*" (*G* 4:462; emphasis added). At the outset, the task was to understand how pure reason could be practical; that is, we wanted to understand how a mere idea (*viz*., the form of universal validity of maxims as laws), which comes from a pure realm devoid of empirical incentives, could generate an interest in us and serve as the determining ground of the will. Kant then ends up thinking that the idea of the intelligible world might itself be the incentive. But this is just another pure idea, and the problem comes full circle. This is the second point I want to draw from Kant's remarks in the *Groundwork*.[196] I call it the **problem of pure reason's practicality**. It comprises several issues. We cannot understand how a human being could take an interest in morality, nor can we comprehend how an idea of pure reason could create feelings of pleasure and displeasure or serve as the determining ground of the will. At the center of this problem is the relation between the intelligible and empirical self. If we could fully comprehend that relation, then we would understand how pure reason could be practical.

[196] Kant reiterates this point in nearly identical terms in the *Critique of Practical Reason*: **(4)** "For, how a law can be of itself and immediately a determining ground of the will (though this is what is essential in all morality) is for human reason an insoluble problem and identical with that of how a free will is possible" (5:72).

Kant underscores these points yet again in the *Critique of Practical Reason*. He emphasizes the connection between the limits on cognition and the incomprehensibility of freedom, and he continues to maintain the material equivalence of our inability to understand freedom and the problem of pure reason's practicality. He does, however, present new arguments in the second *Critique*, and some of these have interesting implications for the incomprehensibility thesis. For my purposes, the most pressing of these questions is whether or not practical reason's extension of our knowledge (which I explained in the last chapter) licenses any theoretical speculation about how freedom is possible. Kant insists that it does not:

> **(5)** But because no intuition, which can only be sensible, can be put under this application, *causa noumenon* with respect to the theoretical use of reason is, though a possible, thinkable concept, nevertheless an empty one. But I do not now claim *to know theoretically* by this concept the constitution of a being *insofar as* it has a *pure* will; it is enough for me to thereby only designate it as such a being and hence only to connect the concept of causality with that of freedom (and with what is inseparable from it, the moral law as its determining ground); and I am certainly authorized to do so by virtue of the pure, not empirical origin of the concept of cause, inasmuch as I consider myself authorized to make no other use of it than with regard to the moral law which determines its reality, that is, only a practical use. (*KpV* 5:55–56).

> **(6)** But is our cognition really extended in this way by pure practical reason, and is what was *transcendent* for speculative reason *immanent* in practical reason? Certainly, but only *for practical purposes.* For we thereby cognize neither the nature of our souls, nor the intelligible world, nor the supreme being as to what they are in themselves . . . But how freedom is even possible and how this kind of causality has to be represented theoretically and positively is not thereby seen; that there is such a causality is only postulated by the moral law and for the sake of it. (*KpV* 5:133)

> **(7)** But this extension of theoretical reason is no extension of speculation, that is, no positive use can now be made of it for *theoretical purposes.* For, since nothing further is accomplished in this by practical reason than that those concepts are real and really have their (possible) objects, but nothing is thereby given us by way of intuition of them (which can also not be demanded), no synthetic proposition is possible by this reality granted them. Hence this disclosure does not help us in the least for speculative purposes, although with respect to the practical use of pure reason it does help us to extend this cognition of ours. (*KpV* 5:134)

Kant thinks that practical reason extends our knowledge by showing that we do indeed possess transcendental freedom, but it does not give us any further information about how this kind of freedom is possible. Our knowledge of freedom is strictly mediated by the moral law. We learn from practical reason *that* we are free, but we do not come to understand *how* we are free. This is the third theme of the incomprehensibility thesis. We have practical knowledge of freedom, but we do not have theoretical knowledge of how it works.[197] I call this the **theoretical knowledge constraint**.

The textual evidence does not end there; it extends far beyond the *Groundwork* and the second *Critique*, showing up in many places before and after the publication of the first *Critique*. Indeed, the incomprehensibility thesis makes appearances in Kant's notes and lectures as early as the 1760s. It is present in some of his earliest notes and lectures on the subject, and some of these provide important clues about his reasoning, further explaining why freedom is beyond our comprehension:

> **(8)** We know well what proceeds from freedom and its presupposition, and it is also necessary for us to presuppose it. But no one can comprehend the origination of a free action, since it is the beginning of all origination. (R 4180)
>
> **(9)** It will still be hard to comprehend by speculative understanding how a derivative being <*ens derivativum*> can perform original acts <*actus originarios*>; but the reason that we cannot comprehend it lies in our understanding, *for we can never conceive the beginning*, but rather only what happens in the series of causes and effects. But the beginning is the boundary of the series, yet freedom makes

[197] I am inclined to go against the grain and call this practical "knowledge," siding with Patrick Kain, Karl Schafer, and others. And (to some extent) this puts me at odds with Andrew Chignell. Chignell divides subjective grounds for assent (*Fürwahrhalten*) into epistemic and non-epistemic grounds, and, since we do not have theoretical reasons for believing that we are free, he lumps freedom in the category of "moral belief" along with belief in God and the afterlife. Unlike knowledge, practical belief (*Glaube*) does not have an objectively sufficient ground, but moral belief is also distinct from other forms of assent — persuasion (*Überredung*) and opinion (*Meinung*) — because moral beliefs have a non-epistemic (practical) subjective ground. See Chignell, "Belief in Kant." Cf. Kain, "Practical Cognition, Intuition, and the Fact of Reason," 217–25 and Schafer, "Practical Cognition and Knowledge of Things in Themselves."

I must concede, however, that Chignell is right that the strict grounds that are required for *Wissen* are not given in the case of freedom, and although Kant occasionally refers to "practical cognition" (*praktischen Erkenntnis* [e.g., B xxi; *KpV* 5:20, 5:57, and 5:103]), he never uses the words '*praktisches Wissen*.' So our rational belief in freedom is akin to knowledge, but it might be appropriate to withhold the objective 'knowledge' label, given Kant's technical definition of the term.

> wholly new divisions for a new beginning; it is on that account difficult to comprehend. But because the possibility of such freedom cannot be comprehended, it does not yet follow from this that, because we cannot comprehend it, there also could not be any freedom. *But freedom is a necessary condition of all our practical actions.* So just as there are other propositions that we do not comprehend but which presuppose a necessary condition, so are we also independent by the concept of transcendental freedom. (*Metaphysik Pölitz*; 28:270)

> **(10)** One can have no insight into the possibility of freedom, because one can have no insight into a first beginning, whether the necessity in existence in general or in freedom in the origination of events. For our understanding cognizes existence through experience, but reason has insight into it if it cognizes it *a priori*, i.e., through grounds (that, namely, which is not necessary in accordance with identity, rather which is posited *realiter*); now there are no grounds for that which is first, thus no insight into it is possible through reason. (R 4338)

These two *Reflexionen* and this excerpt from a 1770s metaphysics lecture refer to something mentioned above in the context of the **cognition constraint**. It was said there that we cannot cognize "absolute causal spontaneity" because a first beginning is incompatible with the understanding's concept of causality, which is a condition of the possibility of experience. What is said in **(9)** underscores this claim. Theoretical cognition of freedom thus faces two barriers: no object is given through intuition, and our empirical concept of causality cannot process a first beginning. In **(10)**, Kant takes this even further as he says not only that we cannot cognize a first beginning through *experience*, he also claims that we cannot cognize it *a priori* through reason since a first beginning has no grounds through which it could be comprehended. Neither experience nor reason can give us any insight into the nature of a spontaneous, first cause. It is inscrutable. This is the fourth and final component of the incomprehensibility thesis: I refer to it as the **first beginning problem**.

Kant continues to bring up the incomprehensibility of freedom throughout his later works. It appears several times in the *Critique of the Power of Judgment*,[198] the *Metaphysics of Morals*,[199] and *Religion within the Boundaries of Mere Reason*.[200] These texts address a variety of different questions, but the four themes of the incomprehensibility thesis continue to resonate: First, we have no theoretical cognition of the intelligible self or the transcendental freedom it possesses. Second, we cannot know how pure reason creates an interest in us, for this would involve determinate knowledge of the intelligible character and its relation to the empirical world. Third, although we have practical cognition through the fact of reason, this knowledge is only "for a practical use." We know we are free because the moral law binds our will, but this practical cognition gives theoretical reason no insight into the mechanism of transcendental freedom. Fourth, we are unable to comprehend a first beginning. A first cause violates the conditions of the possibility of experience, and it cannot be comprehended a priori through reason.

---

[198] Quoting all the relevant passages would involve a great deal of redundancy; I limit myself here to some characteristic excerpts from these later works. In the *Critique of Judgement*, Kant writes **(11)**: "There is thus an unlimited but also inaccessible field for our faculty of cognition as a whole, namely the field of the supersensible, in which we find no territory for ourselves, and thus cannot have on it a domain for theoretical cognition either for the concepts of the understanding or for those of reason, a field that we must certainly occupy with ideas for the sake of the theoretical as well as the practical use of reason, but for which, in relation to the laws from the concept of freedom, we can provide nothing but a practical reality, through which, accordingly, our theoretical cognition is not in the least extended to the supersensible" (*KU* 5:175). Cf. *KU* 5:195, and *KU* 5:275.

[199] He writes **(12)**: "When a human being is conscious of a duty to himself, he views himself, as the subject of duty, under two attributes: first as a *sensible being,* that is, as a human being (a member of one of the animal species), and secondly as an *intelligible being* (not merely as a being that has reason, since reason as a theoretical faculty could well be an attribute of a living corporeal being). The senses cannot attain this latter aspect of a human being; it can be cognized only in morally practical relations, where the incomprehensible property of *freedom* is revealed by the influence of reason on the inner lawgiving will" (*MS* 6:418).

[200] It comes up frequently in *Religion within the Boundaries of Mere Reason* **(13)**: "Thus freedom — a property which is made manifest to the human being through the determination of his power of choice by the unconditional moral law — is no mystery, since cognition of it can be *communicated* to everyone; the ground of this property, which is inscrutable [*unerforschliche*] to us, is however a mystery, since it is *not given* to us in cognition" (*Rel* 6:138). **(14)** "Hence in a practical context (whenever duty is at issue), we understand perfectly well what freedom is; for theoretical purposes, however, as regards the causality of freedom (and equally its nature) we cannot even formulate without contradiction the wish to understand it" (*Rel* 6:145). See also 6:59, 6:117–18, 6:171, and 6:191. Cf. *Metaphysik Mrongrovius* 29:918.

4.2.2. The Argument for the Incomprehensibility Thesis

Having identified these features of the incomprehensibility thesis, I can now formulate the argument for it. For the sake of clarity, I have broken it down into a **main argument** and two **auxiliary arguments.**

The Main Argument

(1) We can comprehend transcendental freedom only if the following conditions are met:
  (a) We can comprehend a spontaneous first cause.
  (b) We can comprehend the relation between a subject's intelligible character and empirical character
(2) Conditions (a) and (b) are not met.
Therefore,
(3) We cannot comprehend transcendental freedom.

Auxiliary Argument (a)

(1) We can comprehend a spontaneous first cause only if we can either cognize it empirically through experience or cognize it a priori through reason
(2) We cannot cognize a spontaneous first cause empirically or a priori.
Therefore,
(3) We cannot comprehend a spontaneous first cause.

Auxiliary Argument (b)

(1) We can comprehend the relation between a subject's intelligible and empirical character only if the following conditions are met:
  (c) We can comprehend how an idea of pure reason (*viz.*, the universal validity of maxims as laws) could create an interest in us.
  (d) We can comprehend how pure reason could create feelings of pleasure or displeasure.
(2) Conditions (c) and (d) are not met.
Therefore,
(3) We cannot comprehend the relation between a subject's intelligible and empirical character.

The first premise of the **main argument** is definitional. Recall from chapter 1 that transcendental freedom is defined negatively, as "independence from everything empirical and so from nature generally" (*KpV* 5:97) and positively as "an absolute causal spontaneity beginning from itself [von selbst] a series of appearances" (A446/ B474). Condition (a) follows clearly enough from the positive definition, but condition (b) requires some explanation.

As we then saw in chapter 2, the key to the resolution of the Third Antinomy — the crucial distinction Kant makes so that freedom and determinism do not conflict with one another — is that transcendental freedom pertains only to an agent's intelligible character qua noumenon/thing in itself. The laws of nature, which deterministically cause an agent's actions, pertain only to her empirical character qua phenomenon/appearance. But this explanation is too thin; we are lacking an account of exactly how the intelligible character determines the empirical. We cannot comprehend how transcendental freedom works unless we comprehend this relation between intelligible and empirical character. Consider, once more, Kant's discussion of the malicious lie. The empirical investigation into what caused the lie will discover only natural, empirical causes in accordance with deterministic laws. But Kant says that we can somehow disregard all of that and consider the lie to be a consequence of the agent's intelligible character:

> Now even if one believes the action to be determined by these causes, one nonetheless blames the agent . . . for one presupposes that it can be entirely set aside how that life was constituted, and that the series of conditions that transpired might not have been, but rather that this deed could be regarded as entirely unconditioned in regard to the previous state, as though with that act the agent had started a series of consequences entirely from himself. This blame is grounded on the law of reason, which regards reason as a cause that, regardless of all the empirical conditions just named, could have and ought to have determined the conduct of the person to be other than it is. And indeed one regards the causality of reason not as a mere concurrence with other causes, but as complete in itself, even if sensuous incentives were not for it but were indeed entirely against it; the action is ascribed to the agent's intelligible character: now, in the moment when he lies, it is entirely his fault; hence reason, regardless of all empirical conditions of the deed, is fully free, and this deed is to be attributed entirely to its failure to act. (A555/ B583)

Transcendental freedom, the capacity to act as a spontaneous first beginning, belongs to the intelligible self, yet it manifests its effects in the empirical world. The malicious lie was something that happened in space and time, but if we are to regard it as free "in the moment when he lies," we must think of it as an effect of agent's intelligible character. Comprehension of this relation is therefore essential to our comprehension of freedom. This, as we saw above, concerns the issue of

how an idea of pure reason can create an interest and how pure reason can cause feelings of pleasure and displeasure. Thus, the soundness of the **main argument** depends entirely on the soundness of the auxiliary arguments.[201]

The first auxiliary argument concerns our inability to comprehend a spontaneous first cause. Once again, there is a necessary condition that fails to obtain, but this time it consists of a disjunction between two modes of cognition. We have only two ways that we could come to comprehend a spontaneous first cause: cognizing it through experience or a priori through reason. But these two means of cognition will be of no help in the case of a first beginning. As it was explained earlier, there are two reasons why experience does not give us any insight into a spontaneous, noumenal cause. First, no object is given through sensible intuition. Second, the concept of causality is capable of connecting only objects of experience in accordance with a causal law. Even if *per impossibile* such noumenal objects of intuition were given, we would not be able to cognize them through the understanding's conceptual schemata. It would be an intuition without a concept, which Kant famously describes as "blind" (A51/ B75). Unlike the understanding, however, reason, does not always require that an object be given. Indeed, one of its distinguishing features is that reason entertains ideas for which no corresponding object of cognition is given (e.g., God, the soul, etc.). But, as we saw, this is precisely what ensnares reason into various traps (the antinomies, paralogisms, and the ideal). Absent any concrete grounds (empirical or a priori), reason has no means by which to cognize anything. It spins its own wheels positing objects where none is given in order to satisfy its demand for completeness. Thus, reason and the understanding cannot provide any insight into the nature of a first beginning because there

[201] More precisely, the soundness of the main argument requires the soundness of only one auxiliary argument. Since both (a) and (b) are necessary conditions, if either one of them is not met, then we cannot comprehend transcendental freedom. Nevertheless, I believe both auxiliary arguments are sound and thus we have two distinct explanations for the incomprehensibility of freedom.

are no grounds through which such a cognition would be possible. The implications of the **cognition constraint** are clearly at work here.

The second auxiliary argument concerns the relation between an agent's intelligible and empirical character. As I argued earlier, Kant claims in both the *Groundwork* and the second *Critique*, that comprehending how freedom is possible is "just the same as" understanding how pure reason could be practical. And this is explicated in turn by conditions (c) and (d). In order to comprehend how pure reason could be practical, we would need to know how a pure idea (respect for the moral law) could create an interest in an empirical being, or alternatively, how a pure idea could create feelings of pleasure and displeasure. As Kant repeatedly emphasizes, the dual character of the human being is precisely what makes this incomprehensible:

> **(15)** If the human being were merely intellectual, then we could have insight into his power of choice through reason; likewise if he were a *brutum*. But not as a sensible and rational being, since his action is subsequently a *phaenomenon*, but antecedently a *noumenon* under practical laws. (R 4788)

The freedom of a purely rational being is relegated entirely to a single realm: the intelligible. It is a rational being exercising its agency according to the dictates of reason. In such cases, there is no question of the relation between an intelligible act of reason and an empirically determined action in space and time. In the case of human agency, things are not so simple, as Kant explains several times in his lectures:

> **(16)** How it is possible to be a cause of appearance without itself being appearance cannot be conceived by any human being, nevertheless, we can conceive from this cause that the mechanism of nature does not conflict with freedom, because as appearance a human being is reckoned to the sensible world, and just the same human being is reckoned as intelligence to the intelligible world. (*Metaphysik Pölitz*, 1790–91, 28:583)

> **(17)** But we can cognize things not in themselves, what and how they are in themselves, but rather only in appearance, therefore we also know a human being only as he is exhibited to us in the form of sensibility as phenomenon, therefore his actions, but not his determining grounds: *it is thus impossible to prove the real*

> *possibility of the absolute self-determination of a human being, or how a human being freely determines himself and nevertheless at the same time is subjected to natural laws*, for that would necessarily require that freedom would be an object of possible intuition. (*Metaphysik Vigilantius* 29:1020)

Again, by relegating freedom and determinism to separate realms, the Third Antinomy demonstrates their compatibility, but it does not provide a concrete account of *how* the intelligible character determines the empirical. Indeed, it must not provide such an account. To do so would overstep the critical boundaries that Kant defends, and he acknowledges as much:

> **(18)** Thus in the judgment of free actions, in regard to their causality, we can get only as far as the intelligible cause, but we cannot get **beyond it**; we can know that actions could be free, i.e., that they could be determined independently of sensibility, and in that way that they could be the sensibly unconditioned condition of appearances. But why the intelligible character gives us exactly these appearances and this empirical character under the circumstances before us, to answer this surpasses every faculty of our reason, indeed it surpasses the authority of our reason even to ask it. (A556–57/ B584–85)

Kant says here that we can know that the malicious lie *could* be free, but we cannot know how the liar's intelligible character gave rise to this empirical character. The mechanism of transcendental freedom is obscure.

Several commentators have made attempts on Kant's behalf to identify the mechanism of transcendental freedom — to explain the relation between intelligible and empirical character. In chapter 2, I also provided such an explanation. My dream machine analogy was an attempt to explain a logically possible relation between indeterministic, free choices and the deterministic, empirical world. But if that is what I was doing in chapter 2, how is it consistent with what I am saying now about the impossibility of comprehending this relation? I could be accused of doing the very thing that Kant says we cannot do. In order to understand why that is not so, recall once more Allen Wood's comparison of the Third Antinomy with the role of a defense attorney. The

story I told (just like those given by Wood, Watkins, Grenberg, Pereboom, etc.)[202] should not be seen as a positive account to which Kant is committed. Wood says that we should think of a positive account of how transcendental freedom works as being akin to a fairy tale that involves mythical creatures but which contains two non-contradictory propositions expressing freedom and determinism respectively:

> It would show that *p* and *q* do not contradict each other. It would not matter in the least to its soundness that the narrative as a whole is false, a pure fiction, talking about fairies, witches, noble-minded handsome princes, and a lot of other things found nowhere in reality and believed in only by contemptibly superstitious people . . . The fact that noumena (and noumenal selves) are mentioned in the story also makes no difference. We are no more committed to their existence than to that of fairies or witches . . . If Kantian ethics is to remain consistent with Kant's own views about the uncognizability of the intelligible world, then it ought to read Kant's story about noumenal freedom in exactly the same way as this fairy tale . . . No doubt this fiction leaves us — as Kant himself says — with no positive comprehension whatever of how freedom and natural causality in fact coexist in the real world.[203]

The task of the Third Antinomy is to show how transcendental idealism could make it possible for freedom and determinism to coexist. Seen in this way, it does not commit us to a positive account of how this works. We cannot know how pure reason has effects in the empirical world, and thus we cannot comprehend how transcendental freedom is possible.

---

[202] Eric Watkins emphasizes Kant's claim that the noumenal world contains the grounds of the world of appearance "*and so too of its laws*" (*G* 4:453). He argues that Kant's view is that free noumenal choices could give rise to the laws of nature in such a way that actions in the empirical world are caused by both a free choice and deterministic laws. See Watkins, *Kant and the Metaphysics of Causality*, 328ff. I favored Watkins's account in chapter 2. Derk Pereboom lends Kant a hand by way of Luis de Molina, suggesting that God could create all and only those transcendentally free agents whose choices "conform to the deterministic laws that God intends for the phenomenal world" ("Transcendental Freedom," 557). Allen Wood suggests (but does not ultimately endorse) the view that timeless noumenal choices could somehow narrow the set of possible worlds leading to the existence of an empirical character that was freely chosen by my noumenal self. Jeanine Grenberg makes use of Kant's "schematism" chapter, proposing that we can understand empirical character as the "schema" of the intelligible character See Grenberg, "In Search of the Phenomenal Face of Freedom."

[203] Wood, *Kantian Ethics*, 137. I find Wood's fairy tale suggestion helpful, but I believe that he overstates the incomprehensibility thesis here. I do not think that we can entirely dispense with the idea of noumenal selves and treat it as fiction. On my view, we cannot know the relation between noumenal and phenomenal selves, but we cannot discard the idea of noumenal selves entirely.

4.3 The Implications of the Incomprehensibility Thesis

The incomprehensibility of freedom has implications for a variety of questions and debates pertaining to Kant's work — too many to exhaustively explore here. I outline a handful of what I take to be some of the most noteworthy implications, highlighting its impact on particular issues within the secondary literature and its significance for Kantian philosophy more broadly. I have called some of these "negative" in the sense that they concern questions or debates where progress is likely to be limited or even impossible because of the incomprehensibility thesis. I have restricted my focus to three such negative implications: the role that moral feeling plays when we act out of respect for the law, the possibility of conversion, and the extent to which freedom can be used as a bargaining chip in the debate over transcendental idealism. In each of these domains, our inability to comprehend freedom has significant consequences and it places limits on how much we can know about these issues. I have dubbed other implications "positive" in cases where recognizing this limitation provides helpful responses to potential problems or objections that could be (and have been) raised against Kant.

4.3.1 Negative Implications

For more than twenty years, there has been a lively debate between "affectivists," who claim that motivation to act out of respect for the moral law requires an affective component — moral feeling — and "intellectualists," who argue that a purely intellectual judgment, without the assistance of any feeling, can motivate an agent to act out of respect for the law.[204] For sympathetic

[204] These names for the views come from Richard McCarty, "Kantian Moral Motivation and the Feeling of Respect." McCarty, who argues for the affectivist view, is somewhat responsible for sparking the debate. On the intellectualist side, Andrews Reath could be seen as the modern standard-bearer (although the intellectualist view was considered the orthodox view for many years). Reath's, "Kant's Theory of Moral Sensibility" came out shortly before McCarty's, and a flurry of articles have been published in the wake of these two influential papers. Reath's article was

readers of Kant's moral philosophy, this poses something of a dilemma. On the one hand, Kant's account requires "that the moral law determine the will immediately" and the "determination of the will" must not be accomplished "by means of a feeling, of whatever kind, that has to be presupposed in order for the law to become a sufficient determining ground" (*KpV* 5:71). This seems to imply that feeling must not play a role in moral motivation. But on the other hand, the idea that intellectual judgment alone can motivate action seems to be at odds with both common sense[205] and many of Kant's comments on the subject. It is not easy to see how this is possible.

But if my argument above is sound and the **problem of pure reason's practicality** is a central component of the incomprehensibility of freedom, then this should come as no surprise. We cannot comprehend how pure reason could be practical nor can we understand how pure reason could have an effect on pleasure and displeasure.[206] It would therefore be quite troubling if this is the problem affectivists and intellectualists are trying to solve. Perhaps the more charitable interpretation would be to assume that this is not what they are trying do.[207]

---

reprinted in *Agency and Autonomy in Kant's Moral Theory* and McCarty's can be found in his book *Kant's Theory of Action*. In favor of the affectivist view, see Ameriks, "Problem of Moral Motivation"; Herrera, "Moral *Triebfeder*"; Grenberg, "Kant's Theory of Action"; and Guyer, *Freedom, Law, and Happiness*. On the intellectualist side, see Henry Allison, *Kant's Theory of Freedom*; Robert Wolff, *The Autonomy of Reason*; and O'Neill, *Acting on Principle*.

There have also been some attempts to find a middle way between the views such as Erica A. Holberg, "The Importance of Pleasure in the Moral for Kant's Ethics." Owen Ware recently argued for a position that is sympathetic to the affectivist stance but which avoids the "conflict of forces" view, which he and others rightly see as a problem for the affectivists. See Ware, "Kant on Moral Sensibility and Moral Motivation."

[205] As McCarty points out, some commentators take this to be a damning criticism of Kant's view ("Kantian Moral Motivation, 422). It would certainly run against the grain of common sense if Kant is claiming that a mere intellectual judgment that we are morally obligated to do something is sufficient for us to do it. To be fair to the intellectualists, however, this is an oversimplified description of the view. Reath's explanation of it is more nuanced. See Reath, *Agency and Autonomy*, 8–28.

[206] These were evident in **(2)** and **(3)**.

[207] This charitable assumption can be hard to maintain at times as commentators sometimes make it sound like this is exactly what they are trying to do. For instance, in a paper on this topic of the moral incentive, Stephen Engstrom writes, "In his chapter on the springs of pure practical reason, Kant undertakes to explicate, in the light of the *Factum* of reason, *how pure reason is practical* in the case of the human being, and more generally in that of a finite subject having a share in this *Factum*" ("The *Triebfeder* of Pure Practical Reason," 93; emphasis added). By my reckoning, it is a mistake to say that Kant's aim in the second *Critique* is to explain "how pure reason is practical" as Kant says many times (e.g., **[2]**, **[3]**, and **[4]**) that we will never comprehend how pure reason is practical.

But that leaves us with a difficult question about what their aim is and how to avoid the conclusion that they are trying to explain how pure reason is practical. One way of sidestepping the issue would be to claim that the things we cannot comprehend all lie upstream, in the mysterious noumenal realm of intelligible beings, and that the locus of the debate over moral feeling is to be found downstream in the comprehensible, empirical world. A move of this sort has been suggested,[208] but it inevitably runs into a problem. If we restrict our focus to the empirical world we would indeed avoid both components of the incomprehensibility thesis: we would not need to say anything about the free agency of the intelligible self or about its relation to empirical character. The problem, however, is that by shifting the discussion to the empirical world, we make it impossible to talk about human frailty — one of the issues at the heart of the debate. Affectivists have argued that frailty is a problem for the intellectualist view insofar as it demonstrates that an intellectual judgment of moral rightness is insufficient to motivate action. In the *Religion*, Kant characterizes this kind of frailty as one of the three degrees of evil (along with impurity and depravity).[209] And he makes it clear that this evil concerns the noumenal self and its propensity to subordinate the moral law to inclinations of self-love in the case of our fundamental moral disposition (*Gesinnung*).[210] This results in yet another dilemma. The debate between affectivists and intellectualists either involves asking questions about the intelligible self's free agency and its

[208] Although references to the incomprehensibility of freedom are scarce in this debate, it has not been completely ignored. Richard McCarty cites part of **(4) —** the beginning of Chapter III of the second *Critique* — where Kant says that explaining "how a law can be of itself and immediately a determining ground of the will (though this is what is essential in all morality) is for human reason an insoluble problem and identical with that of how a free will is possible" (*KpV* 5:72). Rather than attempting to solve this "insoluble problem," Kant says that the task of the chapter is not to show "the ground from which the moral law in itself supplies an incentive but rather what it effects (or, to put it better, must effect) in the mind insofar as it is an incentive" (5:72). McCarty says that his analysis of the phenomenology of respect must therefore avoid the determining ground question (which is insoluble) and focus instead on the psychological effect that the moral law has. He says, "This psychological effect is the motivating feeling of respect for the law. It constitutes the subjective, force-aspect of motivation" (*Kant's Theory of Action*, 177–78).

[209] See *Rel* 6:29ff.

[210] On the *Gesinnung* view, see *Rel* 6:25. And as for frailty being a kind of evil (and thus pertaining to the noumenal realm), see *Rel* 6:32.

relation to empirical character (in which case they are doing something that Kant thinks we should avoid), or they can restrict the discussion to empirical psychology, which would make it impossible to discuss moral agency, frailty, and radical evil in the way that Kant does.[211] But if the incomprehensibility thesis is taken seriously, then perhaps we would avoid many of these questions altogether by heeding Kant's warning that it is impossible to explain "how pure reason could be practical, and all the pains and labors of seeking an explanation of it are lost" (*G* 4:461). This may be one of the most troubling implications of the incomprehensibility of freedom, for a great deal of interesting work has been done on Kant's theory of moral motivation. But if I am right that it involves unanswerable questions about freedom, then the debate is unlikely to ever reach a resolution.[212]

The second negative implication I address concerns the possibility of conversion. Kant claims in the *Religion* that there is a "radical innate evil in human nature" that consists in the fact that the human being "is conscious of the moral law and yet has incorporated into his maxim the (occasional) deviation from it" (*Rel* 6:32). The question then becomes how reform is possible. How can we convert our moral disposition from evil to good? Kant himself is perplexed by the question. Once more, here is what he says about it:

> How it is possible that a naturally evil human being should make himself into a good human being surpasses every concept of ours. For how can an evil tree bear good fruit? But, since by our previous admission a tree which was (in its predisposition) originally good but did bring forth bad fruits, and since the fall from good into evil (if we seriously consider that evil originates from freedom) is no

[211] Patrick Frierson, in his illuminating account of Kant's empirical psychology, mentions this limitation specifically: "Commentators are correct to note that Kant's account of human agency cannot be limited to any empirical account; for Kant, the moral importance of human choice emerges only when humans' empirical character is explained by reference to a free intelligible character as its ground. Thus however detailed an empirical account of human frailty, the evil of such frailty must ultimately be explicable in terms of a practical perspective on choices that does not reduce them to empirical causes but ascribes them to freedom" (*Kant's Empirical Psychology*, 235).

[212] Nevertheless, as I concede in the next section, this is not to say that the entire debate is an idle dispute and that there is nothing valuable to be learned from it. Just as the alchemists' pursuit of the philosophers' stone yielded many significant discoveries in the field of chemistry, so the debate over moral feeling has revealed a number of interesting things about Kant's theory of action, respect for the law, self-conceit, etc.

> more comprehensible than the ascent from evil back to the good, then the possibility of this last cannot be disputed. For, in spite of that fall, the command that we ought to become better human beings still resounds unabated in our souls; consequently, we must also be capable of it . . . (*Rel* 6:44–45).

The agent freely chooses to subordinate the moral law to self-love — a choice that corrupts her fundamental moral disposition. Kant then wonders how it would be possible for this evil tree to bear good fruit. If her disposition is utterly corrupted by radical innate evil, how could it be the origin of a conversion? Kant claims that, even if it is incomprehensible (and indeed it is),[213] conversion must be possible given that we "ought to become better," making this one of several places in the *Religion* where he invokes the "ought implies can" principle.

This has not gotten Kant off the hook with commentators, however. In a recent paper, Samuel Loncar claims that Kant's argument is flawed. In this excerpt from the *Religion*, Kant seems to be claiming that our fall into evil is incomprehensible but actual (*a fortiori* possible), and if our conversion from evil to good is "relevantly parallel," then it too is incomprehensible but possible.[214] The problem, Loncar argues, is that the fall and the conversion are importantly distinct and thus the parallelism argument fails. They look similar on the surface, insofar as the fall involves having a good disposition but then adopting an evil maxim, and conversion involves having an evil disposition and subsequently reforming. But this similarity breaks down upon closer

---

[213] Even putting aside the problem of an evil tree bearing good fruit, the incomprehensibility of atemporal agency is an issue here. The idea of conversion is inextricably tied to that of change. It requires temporal succession: first the agent was bad, then she became good. And Kant acknowledges as much: "But does not the thesis of the innate corruption of the human being with respect to all that is good stand in direct opposition to this restoration through one's own effort? Of course it does, so far as the comprehensibility of, i.e. our *insight* into, its possibility is concerned, or, for that matter, the possibility of anything that must be represented as an event in time (change)" (*Rel* 6:50).

Kant recognizes that we necessarily think of all changes as taking place in time and therefore in accordance with the causal necessity of nature. In a footnote on the same page he calls the freedom involved here "an inscrutable property" and he says yet again that we will never be able to comprehend the relation between freedom and determinism. He bails himself out once more by making use of the "ought implies can" principle. Given my view on the incomprehensibility of freedom, I agree with Guyer that the "ought implies can" bailout is the best move Kant can make when faced with such a problem. See Guyer, "Problems with Freedom," 201–2.

[214] Loncar, "Converting the Kantian Self" 355–58.

analysis. Kant says that our predisposition for good consists in nothing other than our capacity to be good: we are aware of the moral law but we have not yet incorporated it into our fundamental maxim.[215] This is indeed quite unlike conversion which involves reforming a disposition that has *already* adopted an evil maxim. Loncar ultimately concludes that Kant's view of conversion is to be salvaged by recognizing the necessity of "a self-creating subject who preexists subjection to reason and is indeed responsible for subjecting itself to reason."[216] Kant rejected Fichte's self-positing *Ich*, but Loncar thinks Kant's account of radical evil and conversion is pushed toward this very idea even though it lies "beyond the ken of Kant's philosophy."[217]

This suggestion is an interesting one, but it is not the only way out of the bind. The parallel Kant draws between conversion and the fall might not concern their formal structure at all. The relevant parallel might simply be that they are both incomprehensible.[218] Indeed, if my position on the incomprehensibility of freedom is right, then I think we can (and should) avoid asking any questions whatsoever about how conversion is possible. Just as the debate over moral feeling is entangled with the **problem of pure reason's practicality**, the question of conversion cannot get around the **first beginning problem**. Speculation about the possibility of conversion (and about the fall) necessarily involves asking questions about the spontaneous free agency of the intelligible self. We should not expect to find any definitive answers to these questions even if we cannot help but wonder about them. Indeed, Kant sometimes suggests that we cannot resist this speculation; such is the "peculiar fate of reason" (A vii).

---

[215] *Rel* 6:44

[216] Loncar, "Converting the Kantian Self," 364.

[217] *Ibid.*

[218] I am grateful to Ryan Kemp for this response to Loncar. Ryan Kemp, "Kant and the Problem of Moral Conversion."

The last negative implication that I will explore concerns the interpretive debate over transcendental idealism. The dispute centers around the question of how to interpret the distinction between appearances and things in themselves. The resolution to the third antinomy depends crucially on this distinction, and so this issue has been a significant concern for commentators who are interested in making sense of Kant's view of freedom. The intelligible self, the self qua noumenon/thing in itself is free. The empirical self, the self qua phenomenon/appearance, is causally determined by natural laws and thus unfree. Unfortunately, as I mentioned in the preface, there is no consensus on how to take this distinction, and interpretations are widely divergent. On the one hand, there are those who claim that the intelligible and empirical self are identical, that there is only one world — one set of objects.[219] On the other hand, some deny the identity of the noumenal and phenomenal self, asserting the existence of two, ontologically distinct worlds — two sets of objects.[220]

Those familiar with Kant scholarship are likely to be well acquainted with this debate, and it is not my intention to rehash it here or to defend any particular view. My interest is simply to explore the implications that the incomprehensibility of freedom should have on the debate. It is not uncommon for commentators to invoke freedom as a bargaining chip in this debate, arguing that one view or another fails to make sense of transcendental freedom. While I do not believe that

---

[219] There is a divide in the one-world camp between epistemological and metaphysical views. Both sides share the belief that Kant asserts only one set of objects, but they differ on how to interpret the transcendental distinction. According to the epistemological view, Kant's distinction is between two ways of considering objects or two standpoints we could adopt. Versions of this view have been defended by Henry Allison, Graham Bird, Christine Korsgaard, etc. It would be impossible to give an exhaustive list of citations. According to the metaphysical interpretation, the distinction is ontological. For instance, Rae Langton argues that it consists in distinguishing between unknowable intrinsic properties and knowable extrinsic properties. See Langton, *Kantian Humility*. Cf. Lucy Allais, "Kant's One World" and *Manifest Reality*.

[220] This was the traditional reading for many years, going as far back as Jacobi. P.F. Strawson was critical of the view but he attributes something like it to Kant in *Bounds of Sense*, 238. See also Guyer, *Kant and the Claims of Knowledge*. Once again, I cannot give a complete list of references. For a recent defense of the two-object view and a criticism of the metaphysical one-world interpretation, see Nicholas Stang, "The Non-Identity of Appearances and Things in Themselves."

all of these arguments run afoul of the incomprehensibility thesis, some of them do. There is one tactic in particular that should give us pause. The two-standpoint view is sometimes seen as having an advantage over the two-world view on the grounds that the latter either (A) requires (but lacks) a substantive view of how the two worlds relate or (B) presents a substantive view of the relation albeit one that is implausible.[221] Consider the following examples:

> This two-standpoint interpretation has at least two important advantages over two-world accounts. First, it allows one to make use of Kant's insights about freedom from a practical point of view without making substantive metaphysical assumptions. Especially for those primarily interested in Kant as a moral philosopher, this advantage is considerable. Second, *it helps one avoid difficult problems about how different worlds can relate . . .*[222]

> For example, the two-worlds account faces the pressing problem of how to explain the relationship between the two worlds to which it appeals. It seems that the "noumenal I" must bear some relation to some of the actions that take place in the phenomenal world of time and space. Otherwise, the fact that the "noumenal I" is free and its actions undetermined would have no relevance to the actions for which we normally take people to be responsible. Yet it is unclear how anything in the noumenal world, distinct from the phenomenal world of space and time, could "affect" that world. Thus, the two-worlds view saddles us with a metaphysical apparatus that seems to raise as many problems as it solves.[223]

> On the ontological view [of freedom], the question how the two worlds are related is one which, frustratingly, cannot be answered.[224]

---

[221] The three quotes I give here involve (A), but (B) is not uncommon. For instance, Jonathan Bennet takes issue with the idea that his free noumenal choices might determine facts about the empirical world that predate his birth. See Bennett, "Kant's Theory of Freedom." Korsgaard makes a similar point: "Obviously, if we try to picture how Marilyn's freedom is related to the forces that determine her, we must imagine it either inserting itself somewhere into the historical process, or standing behind the laws of nature from which this historical process necessarily follows. And both of these pictures seem crazy" (*Creating the Kingdom of Ends*, 203).

[222] Frierson, "Two Standpoints and the Problem of Moral Anthropology," 84; emphasis added. I would agree with Frierson and Korsgaard that two-standpoint views do not have the difficult problem of explaining how the two worlds relate. The two-standpoint view claims that the incommensurability of the practical and theoretical standpoints allows us to avoid the issue entirely (although I share Eric Watkin's concerns about this move [see *KMC*, 320ff]). My complaint here is that they should not consider this an advantage over the two-world view. They are right to avoid explaining the relation between the intelligible self and empirical self, but they are wrong to think that two-world views should commit to such an explanation.

[223] Dana Nelkin, "Two Standpoints," 565.

[224] Korsgaard, *Kingdom of Ends*, 204.

If I am right about the incomprehensibility thesis, then it is Kant's view that we cannot know the details of the relationship between free noumenal agency and actions in the empirical world. As I see it, the textual evidence on that count is abundant. Thus, if Kant's distinction is indeed one that involves two ontologically discrete worlds, then it is not incumbent on such a view to commit itself to a story of how the two selves relate. Indeed, it must not commit itself to such an account. What should be offered is a logically possible explanation of how there is no contradiction between freedom and determinism: the noumenal self is free and the phenomenal self is determined. Further details about how an agent's intelligible character gives rise to her empirical character are unknowable, and we should neither ask for them nor hold interpretive views responsible for providing them.[225] This is not to say that it is always inappropriate to bring up freedom in the transcendental idealism debate. What would make an objection problematic is if it unfairly requires its opponents to perform the impossible task of explaining the relation between the intelligible and empirical self.

If commentators were to take the implications of the incomprehensibility thesis seriously, the landscape of these three discussions would have to change. The debate over moral feeling would have to situate itself more clearly in relation to the incomprehensibility of freedom by explaining what it is they are trying to understand and how exactly it differs from the inexplicable practicality of pure reason. Open questions about the possibility of conversion from evil would not constitute grounds for finding fault with Kant, as we would have to recognize that the nature of conversion is (and ought to be) mysterious. And participants in the debate over transcendental

[225] Wood puts this point well: "'Just how do freedom and natural causality really relate to each other? What is the metaphysical truth about how they fit together without contradiction? Is the noumenal realm involved in that in any way?' The only permissible Kantian reply to these questions is: 'I do not know, and neither do you, and neither can anyone ever know anything about this'" (Wood, *Kantian Ethics*, 138).

idealism would have to abandon objections that demand an explanation of how the free agency of the intelligible self manifests itself in the empirical world.

4.3.2 Positive Implications

Although the incomprehensibility of freedom limits how much we can know about our free agency, it also opens a space in Kant's practical philosophy for claims that are above the reproach of certain theoretical objections. Some of these include Kant's celebrated moral views that have been admired by generations of scholars in spite of reservations about the connection between these positions and the seemingly untenable the doctrine of transcendental freedom. Here are two familiar claims from Kant's moral philosophy, which he thinks depend crucially on transcendental freedom:

1. The moral law has a special, elevated status: its dictates are categorical and overriding.
2. Human beings, in virtue of their capacity for autonomous action, have an unconditional, incomparable worth/dignity (*Würde*).

In chapters 1 and 3, I presented textual evidence showing that Kant thinks that both of these claims depend on transcendental freedom. The entire structure stands or falls with transcendental freedom. Thus, objections to Kant's view of freedom pose serious problems for his moral philosophy, and if the incomprehensibility thesis can forestall those objections, then it would act as a safeguard for his practical philosophy.

The central points of chapter 3 are also crucial here. As I explained, prescriptions that come from the moral law do not stand on equal footing with those that come from inclinations of self-love. The dictates of morality are given an unparalleled weight that can override all other driving forces, infringing on self-love and striking down self-conceit — nullifying their unwarranted claims on our will (*KpV* 5:73). There is chasm separating autonomy and heteronomy, and it

consists in the fact that they have distinct grounds. For Kant, there are two realms of laws, nature and freedom, and the laws of morality must not come from nature. This is why Kant says that freedom is the *ratio essendi* of the moral law (*KpV* 5:5). Indeed, in a rather remarkable passage from the *Critique of Practical Reason*, Kant says that both the unique respect we have for the moral law and the special status attributed to human beings (in virtue of their capacity for morality) are grounded in transcendental freedom and our membership in the intelligible world. Here is the passage in its entirety:[226]

> *Duty*! Sublime and mighty name that embraces nothing charming or insinuating but requires submission, and yet does not seek to move the will by threatening anything that would arouse natural aversion or terror in the mind but only holds forth a law that of itself finds entry into the mind and yet gains reluctant reverence (though not always obedience), a law before which all inclinations are dumb, even though they secretly work against it; what origin is there worthy of you, and where is to be found the root of your noble descent which proudly rejects all kinship with the inclinations, descent from which is the indispensable condition of that worth which human beings alone can give themselves?
>
> It can be nothing less than what elevates a human being above himself (as a part of the sensible world), what connects him with an order of things that only the understanding can think and that at the same time has under it the whole sensible world and with it the empirically determinable existence of human beings in time and the whole of all ends (which is alone suitable to such unconditional practical laws as the moral). It is nothing other than *personality*, that is, freedom and independence from the mechanism of the whole of nature, regarded nevertheless as also a capacity of a being subject to special laws namely pure practical laws given by his own reason, so that a person as belonging to the sensible world is subject to his own personality insofar as he also belongs to the intelligible world; for, it is then not to be wondered at that a human being, as belonging to both worlds, must regard his own nature in reference to his second and highest vocation only with reverence, and its laws with the highest respect. (*KpV* 5:86–87).

[226] I quoted excerpts of at the conclusion of the last chapter, but it is worth giving the full context. Passages like this one show why I cannot endorse Wood's view. I do not see how he thinks that Kant's ethics can dispense with ideas like noumenal selves, transcendental freedom, etc.

Here and elsewhere, Kant insists that transcendental freedom alone can justify these lofty claims of his practical philosophy.[227] As I have explained, other views of freedom (like Leibniz's) will not suffice.

This adds considerable weight to the criticisms I mentioned in the preface.[228] If Kant's view of transcendental freedom is indeed an incoherent, hopeless failure, then so is his moral philosophy. Without the pure ground of freedom, both the elevated status of human beings and the unique respect for the moral law would be undermined. There would not be anything special about reason after all; it could turn out to be the slave of the passions, just as Hume had argued. An objection of this kind is quite troublesome for those who are seriously interested in Kant's ethics.

One popular way of dealing with this problem is to adopt the two-standpoint view and thus avoid any metaphysical claims concerning the relation of free choices and the empirical world. It is easy to see why this might be attractive. If we were faced with a choice between accepting the two-standpoint interpretation or committing to an absurd metaphysical view of freedom, then the two-standpoint approach would surely be the way to go. I have argued, however, that the incomprehensibility of freedom exposes this as a false dilemma. There is a third alternative. Kant's distinction between things in themselves and appearances might be a metaphysical one, but we do not have to commit ourselves to any particular explanation of how the intelligible character gives rise to its empirical counterpart. Indeed, we should not commit ourselves to such a view. To do so would overstep the boundaries established in the first *Critique*. Kant believes that theoretical reason must be satisfied with mere logical possibility; anything beyond that is unwarranted

---

[227] Kant makes it clear several times that his moral theory requires incomprehensible transcendental freedom. See, for example, *KpV* 5,3–4, 5:97–100. The second epigraph of my dissertation comes from a wonderful expression of this point in Kraus's *Review of Ulrich's Eleutheriology*: "[F]*reedom* insofar as it is the basis of morality *cannot be comprehended*, and insofar as it can be comprehended it cannot serve as the foundation of morality" (8:458).

[228] Kant's view of freedom has been called "incoherent," "metaphysically preposterous," "worthless," and "a hopeless failure." See *supra* note 2, 3, 4, and 5.

speculation. As I argued in chapter 2, this is why Kant is so tentative in his conclusion to the Third Antinomy, claiming that he has not shown the reality or even the real possibility of freedom. When we are confronted with objections about the plausibility of transcendental freedom, Kant's best response would be to concede agnosticism and acknowledge that he does not really have a positive view of how freedom is possible. He does not think we need to do anything further to respond to critics of freedom:

> **(19)** So it is not worth the trouble, either, to refute all the objections leveled against freedom. In this determination of the consciousness of freedom, namely through the categorical imperative, the main question is always the one already illustrated: How is such a categorical imperative possible? This is the most difficult point, since it can neither be proved nor rendered comprehensible; the possibility rests solely on the presupposition of freedom. (27:507)

If Kant is right and we have principled reasons for dismissing criticisms of the plausibility of transcendental freedom, then the incomprehensibility thesis should be seen not merely as a limitation on our knowledge but as a comforting palliative that assuages some of the deepest worries about Kant's critical philosophy. It safeguards not only the two ethical claims mentioned above, it underpins the value of the world itself.[229]

4.4 Conclusion

I would like to conclude by briefly considering objections to the views I have suggested in this paper. I begin with objections to my central argument from section 4.2, and then I address potential reservations about the implications I raised in section 4.3. The first objection concerns

[229] It might be impossible to overstate the importance of freedom in Kant's philosophy. Kant's view is that without freedom, earth would be bereft of value, a meaningless rock drifting through the universe. This line from one of Kant's lectures bears repeating: "If all creatures had a faculty of choice bound by sensuous impulses, the world would have no value; the inner value of the world, the *summum bonum*, is the freedom to act in accordance with a faculty of choice that is not necessitated. Freedom is therefore the inner value of the world" (27:1482). This point is echoed in the third *Critique*: "For (so does everyone judge) if the world consisted entirely of lifeless beings or even in part of living but nonrational beings, then the existence of such a world would have no value at all, because there would exist in it no being that has the slightest concept of a value" (*KU* 5:448–49).

Kant's claims about practical cognition and the extent to which they might undermine the argument I have given for the incomprehensibility thesis. One central component of Kant's conclusion that we cannot comprehend transcendental freedom is his claim from the first *Critique* that we have no theoretical cognition of things in themselves. But in the *Critique of Practical Reason* he seems to walk this back a bit as he suggests that our consciousness of the moral law gives us some insight into the noumenal world, as it demonstrates the reality of transcendental freedom by means of a "practical cognition [*praktische Erkenntnis*]" (*KpV* 5:103).

This objection is not too worrisome, and the reply to it can probably be predicted from my remarks in section 4.2 regarding the **theoretical knowledge constraint**. Kant repeatedly emphasizes that practical cognition demonstrates the reality of freedom without providing any insight into how it works. This is one of the refrains at the end of the second *Critique*: our cognition is indeed extended into the supersensible but "only *for practical purposes*" (*KpV* 5:133). When we think of people in moral terms — making character evaluations and attributing praise or blame — we must put aside the causal necessity of nature and believe that they could have done otherwise. This means that we are justified in asserting the reality of transcendental freedom. But we cannot make the leap from practical reason's certainty to theoretical reason's speculation. As I explained above, Kant's view is that practical reason licenses the assertion of freedom but "how freedom is even possible and how this kind of causality has to be represented theoretically is not thereby seen; that there is such a causality is only postulated by the moral law and for the sake of it" (*KpV* 5:133).[230]

The second objection is somewhat more troubling. I have argued that Kant can legitimately avoid responding to objections to his doctrine of transcendental freedom on the grounds that he

[230] Providing a full account of what practical cognition is and how it differs from theoretical cognition would require its own lengthy discussion. See Karl Schafer "Practical Cognition."

has no positive view of how freedom actually works. The details, whatever they might be, are beyond our comprehension. But I also seem to have committed Kant to a particular metaphysical view of transcendental freedom — one that involves timeless, noumenal agents acting as first causes and serving as the grounds of appearances in the phenomenal world. It looks like I have committed Kant to a metaphysically extravagant view but then made it immune to objection by fiat. I could be accused of claiming both that Kant has no view and that he has a preposterous one. This is likely to come across as exactly the kind of maddening subterfuge I mentioned in the introduction.[231]

My response to this objection is that we must walk a fine line in order to secure a middle way between two untenable positions. On the one hand, there is the view that what Kant says about transcendental freedom should be treated as a total fiction.[232] Every detail he provides about transcendental freedom in his discussion of the Third Antinomy is given as part of a logically possible story, which demonstrates that freedom and determinism are not necessarily incompatible, but Kant is not committed to those details and he does not expect us to accept anything he says there as true. We should pass over transcendental freedom in silence. On the other hand, there is the view that Kant is committed to a particular metaphysical account of transcendental freedom, and he tells us all about it when he resolves the Third Antimony. Neither of these views is particularly attractive.

As palatable as the first interpretation might seem to those who would like to purge Kant of undesirable metaphysical views, it is surely incompatible with what Kant says about the extent of his commitment to transcendental freedom, especially in the second *Critique*. In the Preface,

---

[231] I am grateful to Allen Wood for raising this objection and pressing me on this point.

[232] See Wood, *Kantian Ethics*, 137. Similarly, Graham Bird highlights Kant's talk of intelligible causes as providing a resolution to the antinomy even if they should turn out to be fictional "*erdichtet sein sollte*" (A545/ B573). See *Revolutionary Kant*, 697 and "Antinomies," 238.

Kant tells us that the reality of transcendental freedom is firmly established "taken indeed in that absolute sense in which speculative reason needed it, in its use of the concept of causality, in order to rescue itself from the antinomy in to which it unavoidably falls . . ." (*KpV* 5:3). Using particularly strong epistemic language, Kant says here that our freedom is proven (*bewiesen ist*). So how can Kant's view be that we should pass over transcendental freedom in silence if he tells us that his practical philosophy establishes the reality of precisely that kind of freedom which he had described in the first *Critique*?[233]

If the problem with the first view is that it involves too little of a commitment to transcendental freedom, the problem with the second is that it burdens us with too much of a commitment. Earlier, I mentioned four different ways that commentators have suggested in order to make sense of what Kant says in his resolution of the third antinomy. It could be that free noumenal choices ground the laws of nature.[234] Perhaps God chooses to create noumenal selves whose choices conform to the phenomenal laws.[235] Or it could be something else entirely. Kant's view is that we can never know exactly how this works. I think Wood is right to suggest that we should, in a certain sense, treat these stories as fictions. Their importance lies not in giving us an account of freedom that we are supposed to believe but in showing us that it is at least *logically possible* that transcendental freedom and determinism could be compatible. I think Wood goes too far, however, when he claims that we should dismiss the entire idea of transcendental freedom as a fantasy of transcendent metaphysics.

---

[233] Wood's suggestion is that we should make do with practical freedom. He thinks that transcendental freedom should be "quarantined from Kantian ethics just as strictly as if it carried the plague (*Kantian Ethics*, 138). But I cannot come up with any other way of rendering Kant's remarks in the second *Critique* other than as a firm commitment to the very concept of transcendental freedom that he outlined in the first *Critique*.

[234] Watkins, *Kant and the Metaphysics of Causality*, 328ff.

[235] Pereboom, "Kant on Transcendental Freedom," 557.

This leads back to the question that was at issue throughout this chapter. Exactly how much does Kant think we can know about transcendental freedom? How much of the extravagant metaphysical view does he really expect us to believe? I think the answer is that we must be satisfied with a very thin description of the concept. Kant is committed to the reality of transcendental freedom; so, at the very least we should have some idea how to define that concept (as I did in chapter 1). If we look closely at Kant's various definitions we find a number of recurrent themes, especially the two I highlighted above in the context of the **auxiliary arguments**. Transcendental freedom involves a complete "independence from everything empirical and so from nature generally" (*KpV* 5:97) and the ability to spontaneously "absolute causal spontaneity beginning from itself a series of appearances that runs according to natural laws" (A446 / B474). Thus, it must involve noumenal agency, and the choices of noumenal agents must have effects in the phenomenal world in accordance with deterministic laws of nature. That is all we can know about how transcendental freedom works. We cannot comprehend our noumenal agency (how it converts itself from evil, etc.), and we cannot know how our noumenal agency manifests itself in the empirical world.

The last objection concerns the negative implications I explored in section 4.3. I argued there that the incomprehensibility of freedom has not been fully appreciated in some of the scholarly discussions of moral feeling and conversion. On my interpretation, Kant holds the view that we can never know exactly how pure reason is practical or how conversion is possible. But if Kant holds the view that these are cases of transcendent metaphysical speculation, beyond the bounds of comprehension, then why does he spend so much time discussing them? Chapter 3 of the *Critique of Practical Reason* is dedicated entirely to the incentives of pure reason, and Part

One of the *Religion* is all about the fall of radical evil and the possibility of conversion. It looks like I must accuse Kant of transgressing the very boundaries that he established.

In the case of conversion, I would concede that Kant is going beyond the boundaries of comprehension, but I think he is fully aware of that. He says outright that both the fall from good into evil and the possibility of conversion are incomprehensible (*Rel* 6:45). And when he does speculate about how a conversion might work, he countenances some possibilities but ultimately admits that is a mystery.[236] I think this is in keeping with his views on the incomprehensibility of freedom and the peculiar fate of reason. Moral feeling is a bit more complicated. Not every component of moral feeling is beyond our comprehension. Kant's discussion of it in the second *Critique* begins with a statement of the incomprehensibility thesis.[237] He says we cannot know the ground from which pure reason creates an incentive and that his aim in the chapter is merely to explore what happens downstream from this determination: "what it effects in the mind" (*KpV* 5:72). Much of moral feeling is well within our grasp: we can explore the phenomenology of respect, what it feels like when it strikes down self-conceit, the complicated mix of pleasure and displeasure it involves, and so forth.[238] My concern with the debate in the secondary literature is that it vacillates between the comprehensible and incomprehensible components of pure reason's practicality. I think, just as Kant did, that there is something valuable to be learned from exploring the nature of respect for the moral law, but I believe that discussions of it must be more careful with the boundaries Kant put in place.

---

[236] *Rel* 6:45–53. For a discussion of the movement here and Kant's conclusion that conversion is a mystery, see Ryan Kemp, "Kant and the Problem of Moral Conversion."

[237] I cited it above as **(4)**.

[238] The resemblance between respect and the feeling of the sublime is striking. Kant takes note of this similarity in §27 of the third *Critique*. See *KU* 5:257–60.

The incomprehensibility of freedom could have a variety of such consequences for Kant's views that I have not explored here. But I do not think the negative implications are the principal upshot of the position I have defended. I think the most important result is that Kant offers us a viable way of believing in a kind of freedom that gives unparalleled weight and meaning to our actions. The epistemic limitations that he defends in the *Critique of Pure Reason* safeguard transcendental freedom. The incomprehensibility thesis shields us from objections that threaten to undermine our belief in that kind of freedom — skeptical doubts that would throw the "inner value of the world" into question.

Chapter 5
Conclusion

5.1 Introduction

I have two objectives in this concluding chapter. First, I briefly discuss Kant's third *Critique* to see how it confirms my foregoing conclusions about the incomprehensibility thesis. Most of the arguments and exegesis in the preceding chapters concerned the *Critique of Pure Reason* (first published in 1781), the *Groundwork* (1785), and the *Critique of Practical Reason* (1788). But if I am right about the importance of the incomprehensibility of freedom, then we can reasonably expect it to play an important part in Kant's later texts as well. Although I cited instances of the incomprehensibility thesis from the *Critique of the Power of Judgment* (1790), *Religion within the Boundaries of Mere Reason* (1793), and the *Metaphysics of Morals* (1797), I have not said much about the "architectonic" role that the incomprehensibility of freedom plays in Kant's later work.[239] It is particularly pressing to address the third *Critique* given Kant's description of his aim in the introduction. He says there that he hopes to bridge the "incalculable gulf" between nature and freedom (*KU* 5:175). My goal in section 5.2 is to argue that Kant frames his project in this way precisely because of the incomprehensibility thesis. As I explained in chapter 3, practical reason compels us to believe that we have transcendental freedom, but the epistemic limits described in chapters 2 and 4 prevent us from comprehending how anything in nature is the product of this freedom. I claim that the incomprehensibility thesis, as I have described it, is what gives rise to the gap described in the third *Critique*. In section 5.3, I conclude my project by briefly summarizing the dissertation, offering a synopsis of my view on Kant's multifaceted

[239] This is Kant's grandiose technical term that he uses when talking about the divisions in and systematic unity of his entire philosophical enterprise. It is appropriate here as the problem of the third *Critique* arises out of the disparate treatments of reason's theoretical and practical interests in the first two *Critiques*. See the third chapter of the Doctrine of Method in the first *Critique* (A832 / B860).

concept of freedom. I underscore the unity and coherence of Kant's positions, and I emphasize the importance of the incomprehensibility thesis.

5.2 Confirmation of the Incomprehensibility Thesis in the third Critique

In the introduction to the *Critique of the Power of Judgment*, Kant says that there is a "chasm" (*Kluft*)[240] between "the domain of the concept of nature, as the sensible, and the domain of the concept of freedom, as the supersensible . . ." (*KU* 5:175–76). We might be inclined to wonder what this gap refers to given that Kant had already argued for the compatibility of nature and freedom in the Third Antinomy. Indeed, he tends to boast confidently about his success in eradicating the contradiction: "Thus, freedom and nature, each in its full significance, would both be found in the same actions, simultaneously and without any contradiction, according to whether one compares them with their intelligible or their sensible cause" (A541 / B 569).[241] In this section, I intend to explain what this gap is and what it has to do with the incomprehensibility thesis, paying especially close attention to what Kant says about the highest good.[242]

---

[240] I translate '*Kluft*' as 'gulf,' 'gap,' and 'chasm' throughout the chapter to avoid monotonous repetition.

[241] Cf. *KU* 5:175. Kant says that he "annihilated" the contradiction between nature and freedom in the first *Critique*.

[242] Freedom and morality make several appearances in the first half of the book, which is centered around aesthetic judgments of the beautiful and the sublime, but I have restricted my focus here to the second half of the book, which discusses teleological judgments of nature. My reason for doing this is that aesthetic judgments give us glimpses into the unity of practical and theoretical reason that are slightly less relevant to my aim here. That is not to say, however, that Kant's aesthetics have no bearing on his theory of morality and freedom. By my reckoning, there are at least four connections worth mentioning.

First, as I discussed in chapter 3, moral feeling is similar to experiences of the sublime: they both involve a complex mixture of pleasure and displeasure issuing from reason (see above pp. 137–39). Second, Kant's formalistic account of the pleasure associated with judgments of beauty requires us to take a "disinterested interest" in objects, and they must therefore "please without any interest" (*KU* 5:267). This leads Kant to say that "The beautiful prepares us to love something, even nature, without interest" (*ibid*.) Third, he presents an argument for the conclusion that beauty is the "symbol of morality." See *KU* 5:351–56. And fourth, he claims that "taste is at bottom a faculty for the judging of the sensible rendering of moral ideas (by means of a certain analogy of the reflection on both)" and thus "the development of moral ideas and the cultivation of moral feeling" is the "true propaedeutic for the grounding of taste" (*KU* 5:356). Cf. Ginsborg, "Kant's Aesthetics and Teleology."

I do not mean to downplay the significance of the connections between aesthetics and morality, but I think that Kant's comments about the highest good in the second half of the book have greater relevance for the

In the paragraphs leading up to the "incalculable gulf" passage, Kant says that although it is possible for us to conceive of the compatibility of nature and freedom, we do so by relegating practical and theoretical reason to separate domains.[243] As I argued above, theoretical reason has the resources to explain our cognition of objects in nature, understanding their existence in a causal nexus, but it does not allow us to cognize freedom or comprehend how it brings about effects in the sensible world. Practical reason, on the other hand, confirms the reality of freedom, but it adds nothing to our sensible intuitions or our cognition of objects of appearance. Reason thus appears fragmented when it comes to freedom, and this worry about the unity of reason is what leads Kant to the famous line in the introduction, which I quote here in its full context:

> Now although there is an incalculable gulf fixed between the domain of the concept of nature, as the sensible, and the domain of the concept of freedom, as the supersensible, so that from the former to the latter (thus by means of the theoretical use of reason) no transition is possible, just as if there were so many different worlds, the first of which can have no influence on the second: yet the latter **should** have an influence on the former, namely the concept of freedom should make the end that is imposed by its laws real in the sensible world; and nature must consequently also be able to be conceived in such a way that the lawfulness of its form is at least in agreement with the possibility of the ends that are to be realized in it in accordance with the laws of freedom. — Thus there must still be a ground of the **unity** of the supersensible that grounds nature with that which the concept of freedom contains practically, the concept of which, even if it does not suffice for cognition of it either theoretically or practically, and thus has no proper domain of its own, nevertheless makes possible the transition from the manner of thinking in accordance with the principles of the one to that in accordance with the principles of the other. (*KU* 5:175–76).

---

incomprehensibility thesis, and so I have restricted my focus to that part of the book. Paul Guyer, Ted Cohen, Hannah Ginsborg, and many others have written helpful commentaries on Kant's aesthetic views and their connections with freedom. See Guyer, *Kant and the Claims of Taste*, *Kant's System of Nature and Freedom*, and *Kant and the Experience of Freedom*; Ginsborg, "Kant's Aesthetics and Teleology," *The Role of Taste in Kant's Theory of Cognition*, and "On the Key to Kant's Critique of Taste;" Cohen, "Why Beauty is a Symbol of Morality"; and Michael Rohlf, "The Transition from Nature to Freedom in Kant's Third *Critique*."

Furthermore, by focusing on the highest good and not merely on aesthetics, I avoid the worry that the connection between freedom and nature might consist in nothing more than a handful of analogies or comparisons of feelings. The concern, as Guyer puts it, is that we might make the mistake of thinking that "any unification of the two realms of theoretical and practical thought can only take place in the highly subjective realms of analogy, symbolism, methodological principles, and so on" (*Kant's System of Nature and Freedom*, 279).

[243] As Kant puts it a little later in the introduction: "The concept of freedom determines nothing in regard to the theoretical cognition of nature; the concept of nature likewise determines nothing in regard to the practical laws of freedom: and it is to this extent not possible to throw a bridge from one domain to the other" (*KU* 5:195).

One of the most striking things about what Kant says is that freedom should have some way of realizing its ends in the sensible world. We must be able to conceive of nature in a way that makes room for the ends set by freedom. But, once again, this sounds too similar to the task of the Third Antinomy. Presumably, theoretical reason has already shown how it is logically possible for nature to accommodate the possibility of freedom. Why would Kant take up this project a second time in the third *Critique* and treat it as an unsolved problem? The answer, which becomes clear later in the text, is that Kant has something more specific in mind. It is not enough that there is a logically possible conception of nature that is compatible with freedom; Kant has a particular end of practical reason in mind: the "highest good." This end — the ultimate object of practical reason — must be realizable in the world.

The highest good makes its first appearance in the critical philosophy in the Canon of the *Critique of Pure Reason*. There, Kant asks the three questions that he believes exhaust all the interests of reason: "1. What can I know? 2. What should I do? 3. What may I hope?" (A805 / B833). Taking himself to have adequately explained what we can know in the first *Critique*, Kant turns his focus to the two remaining questions. We have seen in chapter 3 how Kant answers the second question; we should obey the commands of pure practical reason. We should act as the categorical imperative directs us to act. But that is not where the story ends. Practical reason does not reach its terminus with the prescription that we should act only on maxims that we can will to be universal laws. Practical reason's "ultimate end," the highest good, is more grandiose. Kant says that we must hope that happiness will be meted out "in exact proportion with the morality of rational beings" and that "this alone constitutes the highest good of a world into which we must

without exception transpose ourselves" (A814 / B842).[244] And so he restates the third question as "If I do what I should, what may I then hope?" (A805 / B833).

Kant says that this question implicates both practical and theoretical reason. Practical reason tells us what we should do, but happiness concerns the sensible world and thus falls under the domain of theoretical reason.[245] Happiness is a condition that is determined by natural laws and "grounded on empirical principles" (A806 / B834). The question then becomes how there could be any connection between virtue and happiness. If the satisfaction of inclinations is determined empirically by natural laws, how could morality, which concerns our character as noumenal beings, have any impact on the distribution of happiness? Indeed, Kant says that such a connection appears to be impossible: "[B]ecause any practical connection of causes and effects in the world, as a result of the determination of the will, does not depend upon the moral dispositions of the will but upon knowledge of the laws of nature and the physical ability to use them for one's purposes; consequently, no necessary connection of happiness with virtue in the world, adequate to the highest good, can be expected from the most meticulous observance of moral laws" (*KpV* 5:113–14). Kant raises the stakes as he insists that we cannot relinquish this hope lest we forgo morality entirely. He believes that abandoning this necessary object of practical reason (the connection between virtue and happiness) would force us to forfeit the moral law as a phantom of the mind: "Reason sees itself as compelled either to assume such a thing, together with life in such

[244] Although I focus on this facet of the highest good, it is only one of the two components. In addition to the direct proportionality of virtue and happiness, the other component of the highest good is perpetual moral progress. Kant says that pure practical reason requires us to commit to the objective of endlessly striving for moral perfection. Of course, all of our lives end before we reach such a goal, and so Kant posits the immortality of the soul and the existence of God as "postulates of pure practical reason" in the second *Critique.* See *KpV* 5:122–32.

[245] Kant defines happiness as the "satisfaction of inclinations," which experience alone can teach us: "[E]xcept by means of experience I can know neither which inclinations there are that would be satisfied nor what the natural causes are that could satisfy them" (A806 / B834). This means knowledge of happiness involves theoretical reason. That is not to say, however, that practical reason does not also concern itself with happiness. On the contrary, happiness is the principal concern of prudential (though not pure) practical reason.

a world, which we must regard as a future one, or else to regard the moral laws as empty figments of the brain, since without that presupposition their necessary success, which the same reason connects with them, would have to disappear" (A811 / B839).

Kant's solution to this problem in the *Critique of Practical Reason*[246] is to postulate the existence of God and the immortality of the soul. At first glance, this might sound like it is at odds with the epistemic strictures that were meant to banish speculation about these ideas of transcendent metaphysics. But Kant claims that there is no real conflict. The postulates are necessary ideas of practical reason, but they do not license any theoretical knowledge claims about God or the soul.[247] They are valid "only from a practical point of view" (*KpV* 5:133), but they nevertheless make the connection between virtue and happiness a live possibility. If there is an omnibenevolent God, who is the author of the world, and if the soul is immortal, then it is perfectly conceivable that moral rectitude will be rewarded with happiness in the afterlife and we could go on maximizing both for eternity. For the sake of the second *Critique*, Kant believes that this is a satisfactory solution. In the third *Critique*, however, the demand goes beyond this appeal to a "future world" as he argues that reason must find a way of realizing the highest good here and now in the sensible world.[248] For as Kant said in the context of the incalculable gulf, "the concept of freedom should make the end that is imposed by its laws real in the sensible world" (*KU* 5:175).[249]

---

[246] It is also the solution gestured at (but not thoroughly argued for) in the first *Critique*.

[247] The status of the postulates bears a striking resemblance to the point about freedom in chapter 4. But, as noted, freedom stands on slightly firmer ground than God and the soul. See, for example, *KpV* 5:3–4.

[248] There are some, like Guyer, who think that happiness in the sensible world was always Kant's target (even in the first two *Critiques*); he says that Kant never intended to restrict the proportionality of happiness and virtue to the afterlife. But this leads him to dismiss quotations like the one I gave above as instances of Kant being "confused." See *Kant's System of Nature and Freedom*, 317n7. For my argument, it does not matter much when Kant came to the idea. What matters most is that the highest good ultimately requires that virtue be rewarded with happiness here in the sensible world.

[249] Guyer summarizes this nicely: "[T]he fundamental principle of morality does not just constrain our natural ends but itself sets an overarching end for us, the highest good, and this end must be capable of being realized in nature in order for our actions that have it as their end to be rational and coherent . . ." (*Kant's System of Nature and Freedom*, 294).

The incomprehensibility thesis is playing an important role here. Practical reason yields no theoretical insight, and theoretical reason remains agnostic about the inner workings of freedom, particularly the connection between pure practical reason and sensibility. Our inability to comprehend things in nature as products of freedom is precisely the stumbling block that prevents us from connecting the sensible with the supersensible.[250] This might have seemed unproblematic at first, but the absence of a comprehensible connection between nature and freedom now threatens our moral view of the world as Kant believes the highest good hangs in the balance.

Kant's new suggestion in the third *Critique* is that the power of reflective judgment opens up the possibility of a purposive concept of nature that can accommodate this end of practical reason:

> The effect in accordance with the concept of freedom is the final end, which (or its appearance in the sensible world) should exist, for which the condition of its possibility in nature (in the nature of the subject as a sensible being, that is, as a human being) is presupposed. That which presupposes this a priori and without regard to the practical, namely, the power of judgment, provides the mediating concept between the concepts of nature and the concept of freedom, which makes possible the transition from the purely theoretical to the purely practical, from lawfulness in accordance with the former to the final end in accordance with the latter, in the concept of a **purposiveness** of nature; for thereby is the possibility of the final end, which can become actual only in nature and in accord with its laws, cognized. (*KU* 5:196).

Kant's argument for this conclusion about the purposiveness of nature is complex, and I will not reconstruct it here.[251] What is important for my project is the fact that the incomprehensibility of

---

[250] "[T]he word **cause**, when used of the supersensible, signifies only the ground for determining the causality of natural things to an **effect** that is in accord with their own natural laws but yet at the same time is also in unison with the formal principle of the laws of reason, the possibility of which cannot of course be understood" (*KU* 5:195).

[251] The details of the argument lie outside the scope of my interests in this chapter, so I will not retrace all the steps of the argument. But here is a quick summary of its basic movement. Kant thinks our scientific view of the world involves purposive judgments about organisms (e.g., the reciprocal relation between an organism and its parts) He then moves on to claim that not only are organisms judged as purposive but indeed the whole of nature must be directed toward a final end (*Endzweck*). The only kind of being that can serve as a final end is something which is an end in itself (otherwise the process of finding ends would go on ad infinitum). The only such rational being that we are acquainted with is the human being. Thus, Kant concludes that the human being is the final end to which all of

freedom was the driving force behind the problem that motivated the third *Critique*. Insofar as the "incalculable gulf" results from our inability to establish a connection between the use of our freedom and effects in the empirical world, the role of the incomprehensibility thesis is unmistakable. Recall from chapter 4 that one of the important components of the incomprehensibility thesis, **the problem of pure reason's practicality**, was our inability to comprehend how ideas of pure reason could have any effect on feelings of pleasure and displeasure. This is precisely the issue at stake when it comes to the highest good and its realization in the sensible world. If we were able to comprehend how pure reason could have an effect on sensibility, there would be no chasm in the first place. The problem is that we cannot comprehend how this is possible, and this threatens our hope that virtue will be rewarded with happiness.

Framing the project of the third *Critique* in this way confirms the importance of the incomprehensibility thesis, but it might also give rise to a serious objection to the view I have defended. If the gap between nature and freedom is the result of the incomprehensibility thesis, then we might worry that if Kant succeeds at bridging the gap, then he undermines his claims about the incomprehensibility of freedom by demonstrating the connection between freedom and nature. But this is not so. Neither the purposive view of nature nor anything else in the third *Critique* allows us to somehow overcome the incomprehensibility of freedom. Theoretical reason does not gain any insight about the connection between the sensible and supersensible. The principles of teleological judgment must be seen as regulative, not constitutive.[252] These principles do not

---

nature is subordinated. Once again, just like the second *Critique*, this requires us to postulate the existence of God as the moral author of the world.

Guyer writes, "His argument is that the possibility of the highest good as the complete good and ultimate object of morality requires that we conceive of the laws of nature as compatible with the realization of the form of happiness set as a goal by this concept, a possibility that we can conceive only by postulating an intelligent author of the laws of nature who also has an eye on the requirements of the moral law" (*Kant's System of Nature and Freedom*, 314).

[252] See *KU* 5:401. I discussed this distinction in chapter 2. Unlike determining judgments (the contrast class of reflecting judgments), regulative principles about purposiveness do not tell us how nature is *actually* constituted; they only give us a way of thinking about nature, which we cannot dismiss as illegitimate.

produce any theoretical knowledge of freedom or its effects in nature. Kant's theory of reflective judgment removes one of the difficulties that springs from the incomprehensibility thesis; it does not obviate the broader claim. It assuages potential concerns about the possibility of the highest good being realizable in the sensible world. This new view of nature might not have the epistemic status of a constitutive principle, but it does restore our hope that happiness and worthiness of happiness just might coincide in the empirical world after all.

5.3 Summary and Conclusion

I would like to conclude by taking stock of what has been said and offering an overview of the positions I have defended in this dissertation. I began in chapter 1 by providing a novel taxonomy of Kant's various concepts of freedom. Although there are many kinds of freedom at work in Kant's writings, it is important not to overlook their interdependence. At the top of the conceptual pyramid, as I have constructed it, is political freedom. In that context, it has several functions. Most importantly, freedom is the "only original right" belonging to human beings, and it is the proper justification for the existence of the state. Political freedom is then grounded in practical freedom. The reason that human beings have a political right to freedom is because they are self-legislating, rational beings (i.e., they are capable of autonomy). The capacity for moral freedom is what gives human beings a worth that is elevated above everything else in the natural world. On Kant's view, only rational beings have a non-fungible, unassailable value or dignity, and it is this status which undergirds our political rights. At the foundation of autonomy is transcendental freedom, which is a necessary condition of moral freedom (i.e., strong practical freedom). Without transcendental freedom, autonomy would vanish. We would be cogs in the machinery of nature, and, as I explained when discussing the shift in Kant's view over time, he

eventually considered compatibilism to be a contemptible position — one that could not possibly serve as the foundation of his moral philosophy.

In chapter 2, I explained how transcendental freedom first arises in the critical philosophy as a problem for theoretical reason. In its attempt to satisfy reason's insatiable demand for completeness, reason posits a spontaneous, uncaused cause lying mysteriously behind the laws of the empirical world. This gives rise to an ostensible contradiction between freedom and nature, but Kant believes that transcendental idealism mitigates the conflict by locating freedom in the noumenal realm of intelligible beings and determinism in the phenomenal world of empirical ones. Nevertheless, there were several ways that theoretical reason was forced to remain agnostic about transcendental freedom. Kant claimed that he had not demonstrated the reality or metaphysical possibility of freedom. On my view, the resolution of the antinomy did not demonstrate how nature and freedom actually relate, it only sketched a logical possibility of how they *could* be compatible.

In chapter 3, I showed how practical reason "furnishes reality" to the concept of transcendental freedom by means of our practical cognition of the moral law. I explained my view of Kant's "fact of reason" argument, claiming that we have an immediate awareness of categorical prescriptions of how we, as rational agents, ought to act. Not only are we aware of the existence of these oughts, but we recognize them as non-empirical products of pure practical reason. I then put forth a new interpretation of Kant's argument for the existence of transcendental freedom. I claimed that commentators have not adequately appreciated the importance of moral feeling as the missing link between the fact of reason and transcendental freedom. The existence of moral feeling demonstrates that these oughts are not impotent but have effects in the realm of sensibility. This shows that pure reason is (somehow) capable of making a difference in the empirical world.

I presented the central thesis of my dissertation in chapter 4, where I reconstructed Kant's argument for the position that we cannot comprehend how freedom is possible. Having shown in chapter 2 that theoretical reason must remain agnostic about transcendental freedom, I extended that conclusion by showing exactly what it is that we can and cannot know about freedom and what prevents us from acquiring such knowledge. Although practical reason justifies our belief *that* we are free, the critical strictures proscribe the possibility of comprehending *how* we are free. In particular, we cannot know anything determinate about the agency of the noumenal self (e.g., how it converts itself from evil to good), and we cannot comprehend the relation between intelligible and empirical character. I then outlined some of the negative implications of my view, as I believe that certain arguments in the secondary literature are undermined by the incomprehensibility thesis. I concluded on a positive note with the claim that the incomprehensibility of freedom is not merely a limitation: it is a way of defending Kant's view from worrisome objections.

In chapter 5, I offered a cursory look at the aim of the third *Critique* in order to show the continuing relevance of the incomprehensibility thesis. Insofar as the *Critique of Judgment* is meant to reaffirm our hope that virtue will be rewarded with happiness, it is a project that stems from our inability to comprehend how freedom is possible. More specifically, the difficulty arises because theoretical reason is unable to build a bridge between freedom and sensibility (particularly feelings of pleasure and displeasure). Practical reason requires that such a connection exist — its ultimate object demands it. But theoretical is barred from speculating about the mechanism that would make this possible. Kant's theory of reflective judgment opens up a new possibility, as he offers us a teleological view of nature that makes human beings the final end of all creation. This

leads him to posit God as the moral author of the world, and he concludes once again that this is our best hope for the highest good in the world.

The central project of my dissertation was to explain and defend Kant's claim that we cannot comprehend how freedom is possible. In the service of this aim, I explained what freedom means for Kant and how it comes about in the context of both theoretical and practical reason. This junction is the crux of the issue. It all comes down to the fact that these disparate interests of reason must be unified while nevertheless making distinct claims about freedom. Theoretical reason could not establish the reality of transcendental freedom but practical reason required it. Theoretical reason could find no way to connect pure reason with feelings of pleasure and displeasure, but practical reason insisted on the existence of such a connection. Freedom is the single most important intersection of practical and theoretical reason. Hence the famous line from which my dissertation gets its title: "Now, the concept of freedom, insofar as its reality is proved by an apodictic law of practical reason, constitutes the *keystone* of the whole structure of a system of pure reason, even of speculative reason . . ." (*KpV* 5:3–4). Objections to Kant's theory of freedom threaten to crumble the entire edifice of the critical philosophy. The incomprehensibility of freedom is Kant's ingenious way of bolstering the keystone and securing it firmly in place. When critics accuse Kant of putting forth an implausible, incoherent, or worthless theory of freedom, we would do well to remember Kant's claim that he is not in any way committed to a positive view of how freedom is possible. We should rest easy knowing that these critics are tilting at windmills; the cracks in the keystone are merely apparent.

Bibliography

All translations of Kant's writings are taken from *The Cambridge Edition of the Works of Immanuel Kant*. I adopt the standard convention of citing the *Critique of Pure Reason* by referring to the 1781 (A) and 1787 (B) editions. For unpublished notes, I give the *Reflexionen* number. For all other works, I abbreviate the title, and I give the volume and page number from the *Akademie Ausgabe* edition of *Kants Gesammelte Schriften*, edited by *Königlich Preussische Akademie der Wissenschaften* (29 vols. Berlin: de Gruyter, 1900–)

Abbreviations for Kant's Works

A – Anthropology from a Pragmatic Point of View
E – An Answer to the Question: What is Enlightenment?
G – Groundwork of the Metaphysics of Morals
KpV – Critique of Practical Reason
KrV – Critique of Pure Reason
KU – Critique of the Power of Judgment
L_1 – Metaphysik L_1
LJ – Jäsche Logic
MMr – Metaphysics Mrongovius
MS – Metaphysics of Morals
New Elucidations – New Elucidation of the First Principles of Metaphysical Cognition.
Prolegomena – Prolegomena to any Future Metaphysics
R – Reflexionen
Rel – Religion within the Boundaries of Mere Reason
TP – On the Common Saying: That May be Correct in Theory, but it is of No Use in Practice

Allais, Lucy. "Intrinsic Natures: A Critique of Langton on Kant." *Philosophy and Phenomenological Research* 73 (2006): 143–69.

———. "Kant's One World: Interpreting 'Transcendental Idealism.'" *British Journal for the History of Philosophy* 12 (2004): 655–84.

———. *Manifest Reality: Kant's Idealism and his Realism*. Oxford: Oxford University Press, 2015.

Allison, Henry. "Empirical and Intelligible Character in the *Critique of Pure Reason*" in *Kant's Practical Philosophy Reconsidered*, edited by Yirmiyahu Yovel, 1–22. Jerusalem: Springer, 2009.

———. *Essays on Kant*. Oxford: Oxford University Press, 2012.

———. *Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy*. Cambridge ; New York: Cambridge University Press, 1996.

———. "Kant's Concept of the Transcendental Object." *Kant-Studien* 59 (1968): 165–86.

———. “Kant’s Critique of Berkeley.” *Journal of the History of Philosophy* 11 (1973): 43–63.

———. *Kant’s Theory of Freedom*. Cambridge: Cambridge University Press, 1990.

———. *Transcendental Idealism: An Interpretation and Defense*. New Haven, CT: Yale University Press, 2004.

———. “Transcendental Realism, Empirical Realism and Transcendental Idealism.” *Kantian Review* 11 (2006): 1–28.

Ameriks, Karl. “Is Practical Justification in Kant Ultimately Dogmatic?” in *Kant on Practical Justification: Interpretive Essays*, edited by Mark Timmons and Sorin Baiasu, 153–75. New York: Oxford University Press, 2013.

———. “Kant’s Deduction of Freedom and Morality” *Journal of the History of Philosophy* 19 (1981): 53–79.

———. *Kant’s Theory of Mind*. Oxford: Oxford University Press, 2000.

———. “Kant, Hume, and the Problem of Moral Motivation.” *In Kant and the Historical Turn: Philosophy as Critical Interpretation*, 89–107. Oxford: Clarendon Press, 2004.

———. “Kant, Miracles, and *Religion*, Parts One and Two” in *Kant’s* Religion within the Boundaries of Mere Reason: *A Critical Guide*, edited by Gordon Michalson, 137–55. Cambridge: Cambridge University Press, 2014.

Bader, Ralf M. “Kant and the Categories of Freedom.” *British Journal for the History of Philosophy* 17 (2009): 799–820.

Beck, Lewis White. *A Commentary on Kant’s* Critique of Practical Reason. Chicago: Chicago University Press, 1960.

———. *Early German Philosophy*. Cambridge, MA: Harvard University Press, 1969.

———. “Five Concepts of Freedom in Kant” in *Philosophical Analysis and Reconstruction*, *A Festschrift to Stephen Korner*, edited by J.T. Srzednick, 33–51. Dordrecht: Martinus Nijhoff: 1987.

Bennett, Jonathan. *Kant’s Dialectic*. Cambridge: Cambridge University Press, 1974.

———. “Kant’s Theory of Freedom” in *Self and Nature in Kant’s Philosophy*, edited by Allen Wood, 102–112. Ithaca, NY: Cornell University Press, 1984.

Bird, Graham. “The Antinomies: An Entirely Natural Antithetic of Human Reason” in *Kant’s* Critique of Pure Reason: *A Critical Guide*. Edited by James O’Shea. Cambridge: Cambridge University Press, 2017.

———. *The Revolutionary Kant: A Commentary on the* Critique of Pure Reason. Chicago, IL: Open Court Press, 2006.

Bok, Hilary. *Freedom and Responsibility*. Princeton, NJ: Princeton University Press, 1998.

Byrd, Jeremy. "Kant's Compatibilism in the *New Elucidation of the First Principles of Metaphysical Cognition*." *Kant-Studien* 99 (2008): 68–79.

Carnois, Bernard. *The Coherence of Kant's Doctrine of Freedom*. Chicago, Chicago University Press, 1987.

Chignell, Andrew. "Belief in Kant." *Philosophical Review* 116 (2007): 323–60.

———. "Kant, Real Possibility, and the Threat of Spinoza." *Mind* 121 (2012): 635–75.

———. "Rational Hope, Moral Order, and the Revolution of the Will" in *The Divine Order, the Human Order, and the Order of Nature: Historical Perspectives*, 197–218, 2013.

Cohen, Ted. "Why Beauty is a Symbol of Morality" in *Essays in Kant's Aesthetics*, edited by Ted Cohen and Paul Guyer. Chicago: Chicago University Press, 1985.

Delbos, Victor. *La Philosophie Practique de Kant*. Paris: Hachette: 1905.

Descartes, Rene. *The Philosophical Writings of Descartes*. 3 vols. Edited and translated by John Cottingham, Robert Stoothoff, Dugald Murdoch, and Anthony Kenny. (CSM/CSMK).

Engstrom, Stephen. "The *Triebfeder* of Pure Practical Reason" in *Kant's* Critique of Practical Reason: *A Critical Guide*, 90–118, edited by Andrews Reath and Jens Timmerman. Cambridge: Cambridge University Press, 2010.

Esteves, Julio. "The Alleged Incompatibility between the Concepts of Practical Freedom in the Dialectic and in the Canon of the *Critique of Pure Reason.*" *Kant-Studien* 105 (2014): 336–71.

Frierson, Patrick. *Freedom and Anthropology in Kant's Moral Philosophy*. Cambridge: Cambridge University Press, 2003.

———. *Kant's Empirical Psychology*. Cambridge: Cambridge University Press, 2014.

———. "Two Standpoints and the Problem of Moral Anthropology" in *Kant's Moral Metaphysics*, edited by Benjamin J. Bruxvoort Lipscomb and James Krueger, 83–110. Berlin: de Gruyter, 2010.

Ginet, Carl. "Reasons Explanations of Action: Causalist versus Noncausalist Accounts" in *The Oxford Handbook of Free Will*, edited by Robert Kane. Oxford: Oxford University Press, 2005.

Ginsborg, Hannah. "Kant's Aesthetics and Teleology" in *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, 2013. URL=<https://plato.stanford.edu/entries/kant-aesthetics/>

———. “On the Key to Kant’s Critique of Taste.” *Pacific Philosophical Quarterly* 72 (1991): 290–313.

———. *The Role of Taste in Kant’s Theory of Cognition*. London: Routledge, 1990.

Grenberg, Jeanine. “In Search of the Phenomenal Face of Freedom” in *Kant’s Moral Metaphysics*, edited by Benjamin J. Bruxvoort Lipscomb and James Krueger, 111–32. Berlin: de Gruyter, 2010.

———. “Kant’s Theory of Action.” *Mind* 119 (2010): 1198–1205.

Grier, Michelle. “Kant’s Critique of Metaphysics” in *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, 2012. URL =<https://plato.stanford.edu/entries/kant-metaphysics/>

———. *Kant’s Doctrine of Transcendental Illusion*. Cambridge: Cambridge University Press, 2004.

Guéroult, Martial. “Canon de la Raison pure et *Critique de la Raison Pratique.*” *Revue Internationale de Philosophie* 8 (1954): 331–57.

Guyer, Paul. *Kant and the Experience of Freedom: Essays on Aesthetics and Morality*. Cambridge: Cambridge University Press, 2012

———. *Kant and the Claims of Knowledge*. Cambridge: Cambridge University Press, 1987.

———. *Kant and the Claims of Taste*. Cambridge: Cambridge University Press, 1997.

———. *Kant on Freedom, Law, and Happiness*. Cambridge: Cambridge University Press, 2000.

———. *Kant’s System of Nature and Freedom*. Oxford: Oxford University Press, 2005.

———. “Naturalistic and Transcendental Moments” *Inquiry* 50 (2007): 444–64.

———. “Problems with Freedom” in *Kant’s* Groundwork of the Metaphysics of Morals: *A Critical Guide*, 176–202. Cambridge: Cambridge University Press, 2009.

Harper, William and Meerbote, Ralf. *Kant on Causality, Freedom, and Objectivity*. Minneapolis: University of Minnesota Press, 1984.

Herrera, Larry. “Kant on the Moral *Triebfeder*.” *Kant-Studien* 91 (2000): 395–410.

Hobbes, Thomas. *Leviathan*. New York: Penguin: 2017.

Hogan, Desmond. “How to Know Unknowable Things in Themselves.” *Noûs* 43 (2009): 49–63.

Holberg, Erica. “The Importance of Pleasure in the Moral for Kant’s Ethics.” *Southern Journal of Philosophy* 54 (2016): 226–46.

Johnson, Robert. “Kant’s Moral Philosophy” in *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, 2016. URL =<https://plato.stanford.edu/entries/kant-moral/>

Kain, Patrick. “Practical Cognition, Intuition, and the Fact of Reason” in *Kant’s Moral Metaphysics*, edited by Benjamin J. Bruxvoort Lipscomb and James Krueger, 211–32. Berlin: de Gruyter, 2010.

Kant, Immanuel. *An Answer to the Question: What is Enlightenment?* in *Practical Philosophy*, trans. Mary J. Gregor. Cambridge, U.K.: Cambridge University Press, 2008.

———. *Anthropology from a Pragmatic Point of View*, edited and translated by Robert Louden. Cambridge: Cambridge University Press, 2006.

———. *Anthropology, History, and Education*, edited by Günter Zöller and Robert Louden, translated by Mary Gregor, Paul Guyer, Robert Louden, Holly Wilson, Allen Wood, Günter Zöller, and Arnulf Zweig. Cambridge: Cambridge University Press, 2007.

———. *Conjectural Beginning of Human History* in *Anthropology, History, and Education*, edited by Günter Zöller and Robert Louden, translated by Mary Gregor, Paul Guyer, Robert Louden, Holly Wilson, Allen Wood, Günter Zöller, and Arnulf Zweig. Cambridge: Cambridge University Press, 2007.

———. *Correspondence*, translated and edited by Arnulf Zweig. Cambridge: Cambridge University Press, 1999.

———. *Critique of the Power of Judgment*, translated and edited by Paul Guyer and Eric Matthews. Cambridge: Cambridge University Press, 2002.

———. *Critique of Practical Reason* in *Practical Philosophy*, edited and translated by Mary J. Gregor. Cambridge: Cambridge University Press, 2008.

———. *Critique of Pure Reason*, translated and edited by Paul Guyer and Allen Wood. Cambridge: Cambridge University Press, 1998.

———. *Groundwork of the Metaphysics of Morals* in *Practical Philosophy*, trans. Mary J. Gregor. Cambridge, U.K.: Cambridge University Press, 2008.

———. *Lectures on Anthropology*, edited by Allen Wood and Robert Louden. Translated by Robert Clewis, Robert Louden, G. Felicitas Munzel, and Allen Wood. Cambridge: Cambridge University Press, 2012.

———. *Lectures on Ethics*, trans. Peter Heath. Edited by Peter Heath and J.B. Schneewind. Cambridge: Cambridge University Press, 1997.

———. *Lectures on Logic*, translated and edited by J. Michael Young. Cambridge: Cambridge University Press, 1992.

———. *Lectures on Metaphysics*, translated and edited by Karl Ameriks and Steve Naragon. Cambridge: Cambridge University Press, 1997.

———. *Metaphysics of Morals* in *Practical Philosophy*, trans. Mary J. Gregor. Cambridge, U.K.: Cambridge University Press, 2008.

———. *New Elucidation of the First Principles of Metaphysical Cognition* in *Theoretical Philosophy*, *1755–1770*, translated and edited by David Walford. Cambridge: Cambridge University Press, 1992.

———. *Notes and Fragments*, edited by Paul Guyer, translated by Curtis Bowman, Paul Guyer, and Frederick Rauscher. Cambridge: Cambridge University Press, 2005.

———. *Observations on the Feeling of the Beautiful and Sublime and Other Writings*, edited and translated by Patrick Frierson and Paul Guyer. Cambridge: Cambridge University Press, 2011.

———. *On the Common Saying: That may be Correct in Theory, but it is of no use in Practice* in *Practical Philosophy*, trans. Mary J. Gregor. Cambridge, U.K.: Cambridge University Press, 2008.

———. *Prolegomena to any Future Metaphysics* in *Theoretical Philosophy after 1781*, edited by Henry Allison and Peter Heath, translated by Gary Hatfield, Michael Friedman, Henry Allison, and Peter Heath. Cambridge: Cambridge University Press, 2002.

———. *Religion within the bounds of Mere Reason* in *Religion and Rational Theology*, trans. Allen Wood and George di Giovanni. Cambridge: Cambridge University Press, 2001.

———. *Theoretical Philosophy*, *1755–1770*, translated and edited by David Walford. Cambridge: Cambridge University Press, 1992.

———. *Theoretical Philosophy after 1781*, edited by Henry Allison and Peter Heath, translated by Gary Hatfield, Michael Friedman, Henry Allison, and Peter Heath. Cambridge: Cambridge University Press, 2002.

Kemp, Ryan. “Kant and the Problem of Moral Conversion.” (Unpublished).

Kemp Smith, Norman. *A Commentary on Kant’s* Critique of Pure Reason. New York: Palgrave Macmillan: 2003.

Kitcher, Patricia. “Kant on Self-Consciousness.” *The Philosophical Review* 108 (1999): 345–86.

———. “Kant’s Paralogisms” *The Philosophical Review* 91 (1982): 515–47.

———. *Kant’s Thinker*. New York: Oxford University Press, 2011.

———. *Kant’s Transcendental Psychology*. Oxford: Oxford University Press, 1990.

Kitcher, Philip. “Projecting the Order of Nature” in *Kant’s* Critique of Pure Reason: *Critical Essays*, 219–38, edited by Patricia Kitcher. New York: Roman & Littlefield Publishers, 1998.

Kleingeld, Pauline. “Moral Consciousness and the ‘Fact of Reason” in *Kant's* Critique of Practical Reason: *A Critical Guide*, edited by Andrews Reath and Jens Timmerman. Cambridge: Cambridge University Press, 2010.

Kohl, Markus. “Kant on Determinism and the Categorical Imperative.” *Ethics* 125 (2015): 331–56.

———. “Kant on Freedom of Empirical Thought.” *Journal of the History of Philosophy* 53 (2015): 301–26.

———. “Kant on Spontaneity.” *Kantian Review* 20 (2015): 479–83.

———. “Transcendental and Practical Freedom in the *Critique of Pure Reason*.” *Kant-Studien* 105 (2014): 313–35.

Konhardt, Klaus. “Die Unbegreiflichkeit der Freiheit. Überlegungen zu Kants Lehre vom Bösen.” *Zeitschrift für philosophische Forschung* 42 (1988): 397–416

Korsgaard, Christine. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press, 1996.

Langton, Rae. *Kantian Humility: Our Ignorance of Things in Themselves*. Oxford : Oxford University Press, 1998.

———. “Kant’s Phenomena.” Philosophy and Phenomenological Research 73, no. 1 (2006): 170–85.

Leibniz, Gottfried Wilhelm. *Theodicy: Essays on the Goodness of God, the Freedom on Man and the Origin of Evil*, translated by E. M. Huggard. La Salle, IL: Open Court, 1985

Locke, John. *An Essay Concerning Human Understanding*. New York: Penguin, 1998.

Loncar, Samuel. “Converting the Kantian Self.” *Kant-Studien* 104 (2013): 346–66.

Longuenesse, Béatrice. *Hegel’s Critique of Metaphysics*. Cambridge: Cambridge University Press, 2007.

———. “Kant’s Deconstruction of the Principle of Sufficient Reason” in Longuenesse, *Kant on the Human Standpoint*. Cambridge: Cambridge University Press, 2005.

Marshall, Colin. “Kant’s Metaphysics of the Self.” *Philosophers’ Imprint* 10 (2010): 1–21.

———. “Kant’s One Self and the Appearance/Thing-in-Itself Distinction.” *Kant-Studien* 104 (2013): 421–41.

McCann, Hugh. *The Works of Agency: On Human Action, Will, and Freedom*. Ithaca, NY: Cornell University Press, 1998.

McCarty, Richard. "Kantian Moral Motivation and the Feeling of Respect." *Journal of the History of Philosophy* 31 (1993): 421–35

———. *Kant's Theory of Action*. Oxford: Oxford University Press, 2009.

Nauckhoff, Josefine, "Incentives and Interests in Kant's Moral Psychology." *History of Philosophy Quarterly* 20 (2003): 41–60.

Nelkin, Dana K. "Two Standpoints and the Belief in Freedom." *The Journal of Philosophy* 97 (2000): 564–76.

O'Neill, Onora. *Acting on Principle*. Cambridge: Cambridge University Press, 1975.

———. *Constructions of Reason*. Cambridge: Cambridge University Press, 1989.

Paton, H.J. *Kant's Metaphysics of Experience*. London: George Allen and Unwin Limited, 1936.

Pereboom, Derk. "Kant's Metaphysical and Transcendental Deductions" in *A Companion to Kant*, edited by Graham Bird, 154–68. Oxford: Blackwell, 2006.

———. "Kant on Transcendental Freedom." *Philosophy and Phenomenological Research* 73 (2006): 537–67.

Pollock, John. *Contemporary Theories of Knowledge*. New York: Rowman & Littlefield Publishers, 1999.

Rawls, John. *Lectures in the History of Moral Philosophy*. Cambridge, MA: Harvard University Press, 2000.

———. *Political Liberalism*. New York: Columbia University Press, 1993.

———. "Themes in Kant's Moral Philosophy" in *Kant's Transcendental Deductions: The Three* Critiques *and* Opus Postumum, 81–113. edited by Eckart Förster, Stanford: Stanford University Press, 1989

Reath, Andrews. *Agency and Autonomy in Kant's Moral Theory*. Oxford: Oxford University Press, 2006.

———. "Kant's Theory of Moral Sensibility." *Kant-Studien* 80 (1989): 284–302.

Ripstein, Arthur. *Force and Freedom: Kant's Legal and Political Philosophy*. Cambridge, MA: Harvard University Press, 2009.

Robinson, Hoke. "Two Perspectives on Kant's Appearances and Things in Themselves." *Journal of the History of Philosophy* 32 (1994): 411–41.

Roche, Andrew F. “Transcendental Idealism: A Proposal.” *Journal of the History of Philosophy* 51 (2013): 589–615.

Rohlf, Michael. “The Transition from Nature to Freedom in Kant’s Third *Critique*.” *Kant-Studien* 99 (2008): 339–60.

Rousseau, Jean-Jacques. *On the Social Contract*. New York: Hackett, 1988.

Schafer, Karl. “Practical Cognition and Things-in-Themselves” in *The Idea of Freedom: New Essays on the Interpretation and Significance of Kant’s Theory of Freedom*. Edited by Dai Heide and Evan Tiffany. Oxford: Oxford University Press (forthcoming).

Schönecker, Dieter. *Kants Begriff transzendentaler und praktischer Freiheit: Eine entwicklungsgeschichtliche Studie.* Berlin: de Gruyter, 2005.

Sgarbi, Marco. *Kant on Spontaneity*. New York: Continuum, 2012.

Spinoza, Benedictus. *The Collected Writings of Spinoza*, edited and translated by Edwin Curley. Princeton: Princeton University Press, 1985.

Stang, Nicholas. “The Non-Identity of Appearances and Things in Themselves.” *Noûs* 47 (2013): 106–36.

Strawson, P.F. *Bounds of Sense*. New York: Routledge, 1966.

———. “Freedom and Resentment.” *Proceedings of the British Academy* 48 (1962): 1–25.

Sweet, Kristi. *Kant on Practical Life: From Duty to History*. Cambridge: Cambridge University Press, 2015.

Tenenbaum, Sergio. “The Idea of Freedom and Moral Cognition in Groundwork III.” Philosophy and Phenomenological Research 84, no. 3 (2012): 555–589.

Timmons, Mark. Kant’s Metaphysics of Morals: Interpretative Essays. Oxford; New York: Oxford University Press, 2002.

Vaihinger, Hans. *Die Transcendentale Deduktion Der Kategorien (1902)*. Berlin: Kessinger, 2010.

Van Cleve, James. *Problems from Kant*. Oxford: Oxford University Press, 2003.

Velkley, Richard. *Freedom and the End of Reason*. Chicago, University of Chicago Press: 1989.

Vilhauer, Benjamin. “The Scope of Responsibility in Kant’s Theory of Free Will.” *British Journal for the History of Philosophy* 18 (2010): 45–71.

Walker, Ralph. *Kant, the Arguments of the Philosophers*. New York: Routledge, 1999.

Walsh, W.H. *Kant’s Criticism of Metaphysics*. Edinburgh: Edinburgh University Press, 1997.

Ware, Owen. “Kant on Moral Sensibility and Moral Motivation.” *Journal of the History of Philosophy* 52 (2014): 727–46.

———. “Rethinking Kant’s Fact of Reason.” *Philosophers’ Imprint* 14 (2014): 1–21.

Watkins, Eric. *Kant and the Metaphysics of Causality*. Cambridge: Cambridge University Press, 2005 (*KMC*).

———. “The System of Principles” in *The Cambridge Companion to Kant’s* Critique of Pure Reason, 151–67, edited by Paul Guyer. Cambridge: Cambridge University Press, 2010.

Watkins, Eric and Willaschek, Marcus. “Kant’s Account of Cognition.” *Journal of the History of Philosophy* 55 (2017): 83–112.

Willaschek, Marcus. *Praktische Vernunft: Handlungstheorie und Moralbegründung bei Kant.* Stuttgart: J.B. Metzler Verlag, 1992.

Wood, Allen. *Kant*. Malden, MA: Blackwell, 2005.

———. *Kantian Ethics*. Cambridge; New York: Cambridge University Press, 2008.

———. “Kant’s Compatibilism” in *Self and Nature in Kant’s Philosophy*, edited by Allen Wood, 73–101. Ithaca, NY: Cornell University Press, 1984.

———. *Kant’s Ethical Thought*. Cambridge; New York: Cambridge University Press, 1999.

———. *Kant’s Moral Religion*. Ithaca: Cornell University Press, 1970.

Wuerth, Julian. “Paralogisms of Pure Reason” in *The Cambridge Companion to Kant’s* Critique of Pure Reason, 210–44. Cambridge: Cambridge University Press, 2010.